Identification of atopic dermatitis subgroups in children from two longitudinal birth cohorts

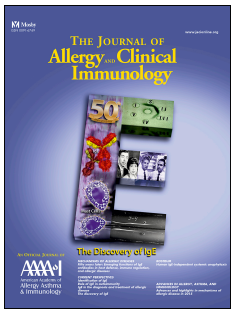
Paternoster, Lavinia; Savenije, Olga E. M.; Heron, Jon; Evans, David M.; Vonk, Judith M.; Brunekreef, Bert

Published in:
Journal of Allergy and Clinical Immunology

DOI:
10.1016/j.jaci.2017.09.044

Publication date:
2018

Accepted Manuscript



Identification of atopic dermatitis subgroups in children from two longitudinal birth cohorts

Lavinia Paternoster, PhD, Olga E.M. Savenije, MD PhD, Jon Heron, PhD, David M. Evans, PhD, Judith M. Vonk, PhD, Bert Brunekreef, PhD, Alet H. Wijga, PhD, A. John Henderson, MD, Gerard H. Koppelman, MD PhD, Sara J. Brown, MD FRCPE

PII: S0091-6749(17)31739-6

DOI: 10.1016/j.jaci.2017.09.044

Reference: YMAI 13102

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 11 June 2017

Revised Date: 7 August 2017

Accepted Date: 20 September 2017

Please cite this article as: Paternoster L, Savenije OEM, Heron J, Evans DM, Vonk JM, Brunekreef B, Wijga AH, Henderson AJ, Koppelman GH, Brown SJ, Identification of atopic dermatitis subgroups in children from two longitudinal birth cohorts, *Journal of Allergy and Clinical Immunology* (2017), doi: 10.1016/j.jaci.2017.09.044.

Identification of atopic dermatitis subgroups in children from two longitudinal birth cohorts

Lavinia Paternoster PhD[1], Olga E.M. Savenije MD PhD [2,3], Jon Heron PhD[4], David M. Evans PhD[1,5], Judith M. Vonk PhD[3,6], Bert Brunekreef PhD[7,8], Alet H. Wijga PhD[9], A. John Henderson MD[4], Gerard H. Koppelman MD PhD[2,3]* and Sara J. Brown MD FRCPE[10,11]*.

*Joint last authors

1. MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, UK
2. University of Groningen, University Medical Center Groningen, Department of Paediatric Pulmonology and Paediatric Allergology, Beatrix Children's Hospital, Groningen, The Netherlands
3. University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, The Netherlands
4. School of Social & Community Medicine, University of Bristol, Bristol, UK
5. The University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Brisbane 4102, Australia
6. University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands
7. Utrecht University, Institute for Risk Assessment Sciences, Utrecht, The Netherlands
8. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
9. National Institute for Public Health and the Environment, Center for Nutrition, Prevention and health Services, Bilthoven, The Netherlands
10. Skin Research Group, School of Medicine, University of Dundee, Dundee, DD1 9SY Scotland, UK
11. Department of Dermatology, Ninewells Hospital, Dundee, DD1 9SY Scotland, UK

Corresponding author:
Dr Lavinia Paternoster
MRC Integrative Epidemiology Unit at the University of Bristol
Bristol Medical School
Oakfield House, Oakfield Grove,
Bristol, UK, BS8 2BN
l.paternoster@bristol.ac.uk

Funding:
This work and LP were supported by a UK Medical Research Council Fellowship (MR/J012165/1). LP & DME work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical Research Council and the University of Bristol (MC_UU_12013/4). DME is supported by an Australian Research Council Future Fellowship (FT130101709). SJB holds a Wellcome Trust Senior Research Fellowship in Clinical Science (106865/Z/15/Z). The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The PIAMA birth cohort was funded by the Netherlands Organization for Health Research and Development, the Netherlands Organization for Scientific Research, the Lung Function of the Netherlands, the Netherlands Ministry of Spatial Planning, Housing, and the Environment, and the Netherlands Ministry of Health, Welfare, and Sport.

Abstract

Background: Atopic dermatitis (AD) is a prevalent disease with variable natural history. Longitudinal birth cohort studies provide an opportunity to define subgroups based on disease trajectories, which may represent different genetic and environmental pathomechanisms.

Objective: To investigate the existence of distinct longitudinal phenotypes of AD and test whether these findings are reproducible in two independent cohorts.

Methods: The presence of AD was examined in two birth cohort studies including 9,894 children from the UK (ALSPAC) and 3,652 from the Netherlands (PIAMA). AD was defined by parental report of a typical itchy and/or flexural rash. Longitudinal latent class analysis was used to investigate patterns of AD from birth to the age of 11 to 16 years. We investigated associations with known AD risk factors, including *FLG* null mutations, 23 other established AD-genetic risk variants and atopic comorbidity.

Results: Six latent classes were identified, representing subphenotypes of AD, with remarkable consistency between the two cohorts. The most prevalent class was early-onset-early-resolving AD, which was associated with male gender. Two classes of persistent disease were identified (early-onset-persistent and early-onset-late-resolving); these were most strongly associated with the AD-genetic risk score as well as personal and parental history of atopic disease. A yet unrecognised class of mid-onset-resolving AD, not associated with *FLG* mutations, but strongly associated with asthma, was identified.

Conclusion: Six classes based on temporal trajectories of rash were consistently identified in two population-based cohorts. The differing risk factor profiles and diverse prognoses demonstrate the potential importance of a stratified medicine approach for AD.

Clinical Implications

Atopic dermatitis ranges from a transient condition to lifelong morbidity. This study has identified distinct subphenotypes of atopic dermatitis in children, which could indicate the importance of a stratified approach to management of this complex disease.

Capsule Summary

Longitudinal latent class analysis models the course of disease subsets over time. Atopic dermatitis in childhood shows diversity in risk factors, prognoses and comorbidities. This study demonstrates robust subphenotypes of AD in two population cohorts.

Key words

Atopic dermatitis
Eczema
Environmental
Genetic
Latent class analysis
PIAMA
ALSPAC

Abbreviations

AD	atopic dermatitis
ALPSAC	Avon Longitudinal Study of Parents and Children
BIC	Bayesian information criterion
FLG	gene encoding filaggrin
LLCA	longitudinal latent class analysis
PIAMA	Prevention and Incidence of Asthma and Mite Allergy

Introduction

In clinical practice atopic dermatitis (AD; eczema) demonstrates a characteristic itchy erythematous rash[1] but it has a heterogeneous presentation with variations in timing of onset, persistence[2,3], distribution, severity, association with allergic sensitisation[4] and comorbidity with other atopic diseases[5,6]. Whilst the classification of eczema cases into atopic and non-atopic forms is commonplace (in part because the underlying aetiology of these may be different[4]), the heterogeneity of longitudinal disease course in AD is less well studied. The majority of AD cases are diagnosed in early childhood and whilst most resolve during childhood, some persist into adulthood. We hypothesised that divergent temporal disease patterns may be caused by different genetic and environmental aetiological mechanisms. Understanding these differences could influence how AD is defined and treated, paving the way for a phenotype-driven, more personalised approach to the management of childhood AD.

AD is a strongly heritable condition. A total of 31 risk loci have been identified in genetic association studies, including 24 loci that were discovered in white European populations[7-13]. The cardinal feature of an itchy erythematous rash is central to all case definitions for AD, but large genetic studies have used a broad case definition, including self-reported AD over a wide age range of paediatric and adult patients. This broad case definition has been necessary to allow collection of the large numbers of cases required for genome-wide analyses, but it does not allow for detailed sub-stratification of AD and dictates that such studies are powered to detect variants common across subtypes of disease, whilst potentially missing variants with more specific effects on subtypes of the disease.

The aim of our study was to investigate the existence of longitudinal subphenotypes of AD and to test whether these findings are reproducible in two independent birth cohorts. We used longitudinal latent class analysis (LLCA), a statistical technique which can be used to model potential subgroups within a dataset to identify different longitudinal patterns of disease. We applied LLCA to cohorts from the United Kingdom and the Netherlands from birth to 16 years or 12 years, respectively. We tested each latent class for association with known genetic and non-genetic risk factors for AD and atopic comorbidities, to investigate the existence of distinct subgroups of disease having different aetiological and prognostic profiles.

Methods

ALSPAC: Avon Longitudinal Study of Parents and Children

ALSPAC is a longitudinal population-based birth cohort study of 14,701 children from Avon, UK. The study protocol has been described previously[14] and further details are in the Online Repository.

Information regarding the presence/absence of rash consistent with AD were extracted from questionnaires completed by the mothers when the children were aged between 6 months and 16.5 years (at 6, 18, 30, 42, 57, 69, 81, 103, 128, 140, 166 and 198 months). At each timepoint AD was defined as a positive response to one of the following questions: "*Has your child had an itchy, dry, oozing or crusted rash on the face, forearms or shins?"* (at 6 months of age); "*Has your child had a skin rash in the joints and creases of their body (e.g. behind the knees, elbows, under the arms) in the past 6-12 months?*" (18-166 months); "*Has your child had an itchy rash which was coming and going for at least 6 months in the past 12 months and confined to the creases of the knees/ankles/elbows or wrists?*" (at 16 years).

Non-genetic risk factors were selected based on existing evidence for association with AD[15] and data availability in the two cohorts. Parental history of AD and asthma were parent-reported in questionnaires completed by the parents or guardians. Breast feeding was coded as a binary variable of never versus any breast feeding as reported by the mother when the child was 15 months old. Cat exposure was coded as a binary variable of 0 or 1+ cats in the home, as reported by the mother at 8 weeks' gestation. Children were classified as asthmatic at 7 and 13 years if a parent answered 'yes' to '*Has your child had asthma in the past 12 months?*'. Total immunoglobulin E (IgE) was measured in venous blood at 7 years and total IgE>75kU/L was defined as elevated.

DNA was obtained from blood and genotypes were determined according to methods described in the Online Repository. Individuals were categorised into two groups: those with and those without any of the 4 *FLG* null mutations tested (i.e. $FLG^{-/-}$ and $FLG^{+/-}$ versus $FLG^{+/+}$)[16]. Genotypes for the remaining 23 established (and replicated) European AD risk variants[13] were combined into a score, with the value representing a sum of the risk alleles carried across the 23 variants.

PIAMA: Prevention and Incidence of Asthma and Mite Allergy

PIAMA is a Dutch multicenter birth cohort of 3963 children from allergic and non-allergic mothers. The study protocol has been described previously[17] and further details are in the Online Repository.

The International Study of Asthma and Allergies in Childhood–based questionnaires were used to report symptoms of AD between 3 months and 11 years of age (3, 12, 24, 36, 48, 60, 72, 84, 96, 132 months). At each timepoint AD was defined as a positive response to both of the following two questions: *"Has your child had an itching rash that was variably present in the last 12 months?"* (or *ever* at 3 and 12 months) and, "*Was this rash present around the eyes/ears, foreside ankles, inner side knees or inner side elbows?"* (also *neck* at 3 and 12 months).

Parental history of asthma was taken from questionnaires which asked "*Have you ever had asthma*?" Any versus never breast feeding was assessed by questions on infant feeding in the questionnaires administered at 3 months and at 1 year of age. Cat exposure was coded as a binary

variable of 0 or 1+ cats in the home at 3 months after birth. Asthma at 7 and 11 years was defined as a parental report of a doctor's diagnosis of asthma at any time and a parental report of asthma in the last 12 months at age 7 and 11. Total IgE was measured in venous blood at age 8 years and >75kU/L was defined as elevated.

DNA was obtained from blood and mouth swabs and genotypes were determined according to methods described in the Online Repository. *FLG* genotype categorisation and the non-*FLG* genetic risk score were constructed, as for ALSPAC.

Statistical analysis

Longitudinal latent class analysis (LLCA) was used to investigate heterogeneity in patterns of AD. As the name suggests, this method is applied in longitudinal settings[18,19], where the aim is to identify distinct subgroups in longitudinal multivariate categorical data. This is akin to cluster analysis, but is more appropriate for binary data and allows for assignment based on probability, rather than definitive partitioning of individuals into classes. Starting with a single latent class, additional classes are added until measures that estimate model fit are optimised. Several statistical criteria (including low Bayesian Information Criterion (BIC), Vuong-Lo-Mendell-Rubin likelihood ratio test and entropy) were assessed to determine the optimal number of classes; full details are given in the Online Repository. Model fitting was carried out in Mplus version 7.0[20].

Model fit was primarily assessed using only those individuals for whom there was no missing AD symptom data. However, to optimise the use of available data and maximise the cohort size, results were compared to analyses which included individuals for whom data were available for ≥50% of the time points studied (≥6 of the 12 time-points in ALSPAC and ≥5 of the 10 time points in the PIAMA cohort). Association analyses primarily focused on this larger (although incomplete) dataset, but results were compared with models from the smaller but more complete dataset.

Associations of risk factors and comorbidities with the latent classes were tested using a manual implementation of the bias-adjusted three step analysis[21]. This method accounts for uncertainty in class assignment (see Online Repository). Associations with established risk factors (sex; family history of atopy; breastfeeding; presence of pet cat in the household; *FLG* loss-of function mutation; genetic risk score of 23 established white European AD variants) were tested using multinomial regression, whilst atopic comorbidities (asthma at ages 7 and 11 or 13 years; elevated IgE at age 7 or 8 years) were tested using logistic regression.

Results

LLCA in the ALSPAC cohort

The prevalence of AD in ALSPAC declined over time (Fig 1a) from 27% in the first year of life to 7% at 16.5 years of age. Data were available from all 12 time points for a total of 3,480 individuals. The 6-class model was considered the best fit to the data (as defined by the lowest Bayesian information criterion), however only small improvements were seen between the 4-class and the 6-class models (Table E1 & Fig E1, Online Repository). We present the results of the 6-class model as the primary analysis, but show the results for the simpler 4-class model in the Online Repository, which for most analyses produced very similar results.

9,894 individuals had data available from at least 6 of the 12 time points and the model fit parameters were broadly consistent with the smaller but more complete dataset (Table E1, Online Repository). Comparison of models from the larger incomplete and smaller but complete datasets showed that the prevalence patterns of AD by class were very similar (Table E3, Online Repository) and only 3% of children (116/3480) changed best-fit class between the 6-class models in each analysis (Table E2, Online Repository).

Fig 2a shows the estimated prevalence of a rash characteristic of AD at each time-point across the 6 classes in the analysis of 9,894 individuals. Descriptions of the classes alongside the labels we gave each class are given in Table I.

The estimated prevalence of a rash characteristic of AD at each time point for the 4-class model are displayed in Fig E3 (Online Repository). The 4 classes can be described as: Unaffected individual or transient AD (61.9%); early-onset-persistent AD (10.7%); early-onset AD resolving by 11 years of age (16.5%); later-onset AD after 3.5 years of age (10.9%). These 4 classes show substantial overlap with the 6-class assignment (Table E4, Online Repository).

LLCA in the PIAMA cohort

The prevalence of AD in PIAMA declined only slightly from 18% in the first year of life to 14% by 11 years of age (Fig 1b). Data were available from all 10 time points for 2,063 individuals. 3,652 individuals had data available from at least 5 of these time points, and we present the results from the analysis of this larger incomplete dataset. LLCA model fit was similar to ALSPAC, with lowest BICs achieved between the 4-class and 6-class models (Table E5, Fig E2, Online Repository), the resulting class patterns following a remarkably similar pattern to ALSPAC (Fig 2b, Fig E4, Online Repository), with comparable class prevalences (Table I, Fig1b).

As for ALSPAC, we present association results for the 6-class model as the primary analysis because this showed best fit. Comparison of assignment between 4- and 6-class models is shown in Table E7 (Online Repository) and association results from the 4-class model are also shown in the Online Repository.

Associations between latent classes with family history and selected environmental risk factors

The associations of 6 classes with potential AD risk factors are summarised in Table II. The results from the smaller but complete dataset and the 4-class models are shown in Tables E9 & E10 (Online

Repository). Similar conclusions could be drawn from these models, unless otherwise specified.

In ALSPAC, taking the 'unaffected or transient AD' class as the baseline category, being female was a risk factor for the early-onset-persistent, mid-onset and late-onset classes, the strongest association being with the late-onset class (OR=1.90 (95% confidence interval 1.48-2.44) $p=4x10^{-7}$). However, male gender was a risk factor for the early-onset-early-resolving class (OR=1.33 (95% CI 1.10-1.61) p=0.004). A similar pattern was observed in PIAMA, where the strongest association with female gender was observed with the late-onset group (OR=1.87 (95% CI 1.21-2.90) p=0.005) and there was evidence of an association between male gender and early-onset-early-resolving class.

Maternal history of AD was associated with all classes in ALSPAC, with the strongest association in the persistent class (OR=3.16 (95% CI 2.60-3.83) $p=4x10^{-31}$). A similar pattern (albeit with weaker evidence for all classes) was observed in ALSPAC for maternal history of asthma, where again the strongest association was with the persistent class (OR=1.54 (95% CI 1.22-1.95) $p=3x10^{-4}$). Paternal history of asthma showed a similar association with this class (OR=1.59), but the smaller sample size meant there was less evidence for this result (p=0.139). Paternal asthma also showed association with the early-onset-late-resolving class (OR=2.53 (95% CI 1.30-4.91) p=0.006). In PIAMA the associations with maternal and paternal history of asthma were similar, with strong associations with the persistent and early-onset-late-resolving groups for maternal history (OR=1.94 (95% CI 1.11-3.40) p=0.021, OR=3.14 (95%CI 1.76-5.61) $p=1x10^{-4}$, respectively) and with the persistent group for paternal history (OR=2.69 (95% CI 1.66-4.36) $p=6x10^{-5}$).

In ALSPAC, breastfeeding was associated with a higher risk of persistent and early-onset-late-resolving AD (OR=1.42 (95% CI 1.11-1.81) p=0.006; OR=1.53 (95% CI 1.12-2.08) p=0.008, respectively). There was little evidence of association with mid- or late-onset classes. In PIAMA, there was little evidence for breastfeeding being associated with any class.

Early-life exposure to a pet cat was not associated with any of the latent classes in the primary analyses for ALSPAC or PIAMA. However this was the only risk-factor in which a difference was seen in the complete-case results in ALSPAC, where there was some evidence of a protective effect of early-life cat exposure on the early-onset-early-resolving class only (OR=0.64 (95% CI 0.46-0.90) p=0.010, Table E9, Online Repository). The same direction of effect was observed in PIAMA but with a weaker and less precise estimate (OR=0.72 (95% CI 0.50-1.04) p=0.081).

Associations between latent classes and atopic traits and comorbidities

The associations of AD classes with elevated total IgE and asthma are displayed in Table III. Raised IgE was associated with the AD classes showing prevalent disease at the time of testing, i.e. 7-8 years of age (the persistent, early-onset-late-resolving and mid-onset classes in ALSPAC and the persistent class in PIAMA).

In ALSPAC, all classes showed association with asthma at 7 and 13 years of age. The associations were strongest for the persistent class (7 years OR=4.94, $p=7x10^{-51}$; 13 years OR=6.04 $p=3x10^{-56}$) in which 29% reported asthma at 7 years of age (compared to 8% of the normal/transiently affected class), increasing to 31% at 13 years (compared to 7% of the normal/transient class). The early-onset-early-resolving class showed the smallest increased risk of asthma at 7 and 13 years of age (ORs 1.52 and 1.72, respectively). In PIAMA the persistent and early-onset-late-resolving group

showed association with asthma at age 7 (persistent OR=9.86 $p=3x10^{-10}$). At age 11-12 years all but the mid-onset-resolving group were associated, again the strongest association being with the persistent group (OR=9.61 $p=2x10^{-9}$).

Associations between latent classes and genetic risk variants

In ALSPAC, all classes other than the mid-onset class showed association with *FLG* null mutations (Table IV). The strongest association was for the persistent group (OR=4.31 (95% CI 3.29-5.63) $p=2x10^{-26}$); the other associated classes had ORs of about half this (2.14-2.30). In PIAMA only the early-onset-late-resolving class was associated with *FLG* null mutations (OR=5.63 (95% CI 2.65-11.95) $p=7x10^{-6}$), however the approximate number of $FLG^{-/+}$ or $FLG^{-/-}$ individuals in the PIAMA analyses were very low (between 7 and 14 individuals per class), so power was limited to identify associations.

The combined genetic risk score encompassing all other AD variants was associated with all but the early-onset-early-resolving and the late-onset classes in the 6-class model in ALSPAC. The association was strongest with the persistent class (OR=1.17 (95% CI 1.12-1.22) for each additional risk allele, $p=2x10^{-13}$). A similar pattern was observed in PIAMA, with the persistent class showing the strongest association and an almost identical effect size to that seen in ALSPAC (OR=1.17 (95% CI 1.07-1.28) $p=5x10^{-4}$).

The associations for individual AD risk SNPs are shown in Table E11 (Online Repository). These analyses are not well powered and should be interpreted with caution, but some patterns are noteworthy. Most variants had the strongest effects in the persistent class and three variants showed consistent associations in ALSPAC and PIAMA: These were rs17881320 in *STAT3*, rs479844 near *OVOL1* and rs6010620 in *RTEL1*. One variant (rs1057258) showed evidence in ALSPAC for association in the opposite direction to that reported previously for AD with the late-onset and early-onset-early-resolving classes (OR=0.73 (95% CI 0.57-0.93) p=0.011; OR=0.80 (95% CI 0.65-0.99) p=0.039, respectively). A consistent direction of effect (though with weak statistical evidence) was observed for the late-onset class and this SNP in PIAMA (OR=0.74 (95% CI 0.45-1.20) p=0.218).

Discussion

Our results provide novel insights into the heterogeneity of AD in childhood. We report six latent classes, representing subphenotypes of AD with remarkable consistency between two independent cohorts. The most prevalent class was early-onset-early-resolving AD (13-15%), which was associated with male gender. This class has a favourable prognosis and is only very weakly associated with asthma in later life. Two classes of persistent disease were identified (early-onset-persistent and early-onset-late-resolving); these were most strongly associated with an AD genetic risk score as well as personal and parental history of atopic disease. Importantly, these classes display strong comorbidity with asthma. A yet unrecognised class of mid-onset-resolving AD, not associated with *FLG* mutations, but strongly associated with asthma, was described. In this class AD prevalence increases sharply from 2.5 years of age peaking at ~6 years. The aetiological factors in this class remain unclear because the subgroup was not strongly associated with many of the known risk factors, but does show strong association with asthma comorbidity.

The clinical application of this LLCA is based on the clear demonstration of distinct classes of AD phenotype with different disease trajectories. The substantial diversity of disease which is defined as 'AD' (or 'eczema') has long been recognised, and clearer subdivisions are an essential prerequisite for the development of stratified medicine approaches which will be needed for the optimal application of novel biological therapies in the more severe subgroups of AD. Therefore, further studies are needed to define the most appropriate combinations of biomarkers and risk factors to detect these subgroups prospectively and at an early age.

There was some evidence of differential strength and presence of associations with risk factors and comorbidities between the classes. The early-onset-persistent class showed the strongest associations (compared to other classes) with the majority of well-established risk factors and markers of severe atopic phenotype, including *FLG* null mutations and a genetic risk score of other AD-associated variants, co-existent asthma and elevated IgE and parental history of atopic disease. The associations with asthma at ages 7 and 11-13 years were strongest with the persistent class, but all AD classes showed evidence of some increased risk of asthma at these ages. Our data did not support the presence of a specific trajectory from AD to asthma (the so-called atopic march), which is in keeping with a previous report from the MAAS study and earlier analyses in ALSPAC[22]. The associations observed with elevated total IgE were most marked during active and persistent disease, in keeping with previous reports[23]. Within the class of early onset disease that resolved before the time of IgE measurement, a smaller proportion of individuals had IgE levels above the threshold defined as 'elevated', in comparison with the class of early onset disease that was still active. Further investigation with earlier IgE measures are required to explore whether such individuals would have had raised IgE at the time of active disease.

Whilst being female was more strongly associated with the early-onset persistent and late-onset classes, there was some evidence that being male was differentially associated with early-onset-resolving classes. It is tempting to speculate that the late-onset class might represent AD induced by behavioural changes in the adolescent child (including increased bathing/showering and the use of fragranced products) which might differ between males and females, but this hypothesis remains to be tested. The male preponderance in AD cases ascertained in infancy has previously been reported[24-27] but the mechanisms accounting for this gender difference are unknown.

There is conflicting epidemiological evidence indicating that breastfeeding may be either a risk factor or protective factor in the aetiology of AD[28] and our analyses have not been able to add clarity to this important question. In the ALSPAC cohort, breastfeeding was associated with the two classes of most long-lasting disease (early-onset-persistent and early-onset-late-resolving AD) but in the PIAMA cohort there was little evidence of breastfeeding being associated with any of the latent classes. This apparent difference could be explained in several different ways: It may be stochastic, (given that all 95%CIs overlap between the two cohorts); it may be a result of the substantially higher prevalence of breastfeeding in the Dutch population compared with the UK; or it may result from reverse causation in the ALSPAC cohort, if, for example, mothers with a strong history of atopic disease are more likely to breastfeed their infants.

We found little evidence for early-life exposure to a pet cat being associated with any of the classes in the main analysis. However, in the complete-case dataset in ALSPAC there was some evidence that cat exposure may have a protective effect on early-onset-early-resolving AD, which is somewhat at odds with previous reports of cat exposure increasing the risk of AD[29,30]. This may indicate a specific beneficial effect of early cat exposure on this more transient phenotype and warrants further investigation.

There was evidence that *FLG* null mutations were associated with all classes, however, as reported previously[31,32], the association was strongest in the group with early-onset-persistent disease. The genetic risk score of the other established AD variants showed a similar pattern, whereby the association was strongest for the early-onset-persistent class, with a striking increase in burden of risk of approximately 17% per additional risk allele. Of note, three variants showed consistent patterns of effects across both cohorts, with stronger associations in the early-onset persistent group, and weaker associations with the other classes. The functional mechanisms of these loci have not been fully defined, but rs17881320 is within *STAT3* (encoding a signal transducer and activator of transcription, an acute phase response protein); rs479844 is near to *OVOL1* (which encodes zinc-finger containing transcription factor); and rs6010620 is within *RTEL1* (regulator of telomere elongation helicase 1). This heterogeneity of effect of genetic variants on different disease profiles, emphasises the need for patient stratification in future genetic studies. Stratification may be used to increase the power to detect variants associated with specific classes; stratification could also allow the identification of phenotype-specific mechanistic pathways as future therapeutic targets.

The similarity and high frequency of AD symptom ascertainment in ALSPAC and PIAMA are strengths of our study. The phenotype definitions used within the cohorts comprised prospective questions to capture diagnostic features of eczema including the changing distribution of skin involvement from infancy to later childhood. One key difference is that PIAMA had shorter follow-up (11 years versus 16 years), which could have limited the ability to detect classes with differences at later ages. Despite this, the class patterns are remarkably similar between ALSPAC and PIAMA (Figure 2). However, neither cohort allowed for investigation of variable AD patterns in adulthood, more subtle AD patterns (such as the transient cases which were indistinguishable from the unaffected individuals in our study) nor in people of ancestries other than white European. We also studied only a limited number of environmental factors and so this work could be extended to investigate the association between latent classes and other potential risk factors.

Although individuals are not assigned to classes with complete certainty (Table E8), the LLCA 3-step

method models this uncertainty and allows for inclusion of individuals with incomplete data to maximise sample size and minimise any loss to follow up bias. We note that our analysis does not formally test for causal direction and some risk factors studied do not entirely precede the onset of disease. Therefore, although the observational associations are interesting, further work should be conducted to investigate causality. A further challenge is that by stratifying AD into smaller subphenotypes of disease, we inevitably lower the power of association testing. Few studies have such detailed longitudinal data and so in order to increase sample sizes in future studies, it will be necessary to extrapolate these data-driven phenotypes into settings where less detailed data are available, such as large data registries[33].

In conclusion, we have identified longitudinal subgroups of AD which have both shared and distinctly different risk factor profiles. Future studies of the aetiology and treatment of this complex trait should take these subgroups of disease into account and in turn this may offer valuable stratified medicine approaches to refining prognostic predictions and therapeutic strategies.

409 Acknowledgements

410 ALSPAC: We are extremely grateful to all the families who took part in this study, the midwives for
411 their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer
412 and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists
413 and nurses. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome
414 Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

415 PIAMA: The authors thank all the children and their parents for their cooperation. The authors also
416 thank all the field workers and laboratory personnel involved for their efforts, and Marjan Tewis for
417 data management.

418 This publication is the work of the authors and LP will serve as guarantor for the contents of this
419 paper.

420

421 References

422 1. Williams H.C. Clinical practice. Atopic dermatitis. *N Engl J Med* **352**, 2314-24 (2005).
423 2. Kim J.P., Chao L.X., Simpson E.L. & Silverberg J.I. Persistence of atopic dermatitis (AD): A
424 systematic review and meta-analysis. *J Am Acad Dermatol* **75**, 681-687 e11 (2016).
425 3. Mortz C.G., Andersen K.E., Dellgren C., Barington T. & Bindslev-Jensen C. Atopic dermatitis
426 from adolescence to adulthood in the TOACS cohort: prevalence, persistence and
427 comorbidities. *Allergy* **70**, 836-45 (2015).
428 4. Johansson E.K., Ballardini N., Bergstrom A., Kull I. & Wahlgren C.F. Atopic and nonatopic
429 eczema in adolescence: is there a difference? *Br J Dermatol* **173**, 962-8 (2015).
430 5. Garcia-Aymerich J., Benet M., Saeys Y., Pinart M., Basagana X., Smit H.A. *et al.* Phenotyping
431 asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic
432 comorbidity cluster. *Allergy* **70**, 973-84 (2015).
433 6. Belgrave D.C., Granell R., Simpson A., Guiver J., Bishop C., Buchan I. *et al.* Developmental
434 profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS*
435 *Med* **11**, e1001748 (2014).
436 7. Paternoster L., Standl M., Chen S.M. Ramasamy A., Bonnelykke K., Duijts L. *et al.* Meta-
437 analysis of genome-wide association studies identifies three new risk loci for atopic
438 dermatitis. *Nature genetics* **44**, 187-192 (2011).
439 8. Ellinghaus D., Baurecht H., Esparza-Gordillo J., Rodriguez E., Matanovic A., Marenholz I. *et al.*
440 High-density genotyping study identifies four new susceptibility loci for atopic dermatitis.
441 *Nat Genet* **45**, 808-12 (2013).
442 9. Weidinger S., Willis-Owen S.A., Kamatani Y., Baurecht H., Morar N., Liang L. *et al.* A genome-
443 wide association study of atopic dermatitis identifies loci with overlapping effects on asthma
444 and psoriasis. *Hum Mol Genet* **22**, 4841-56 (2013).
445 10. Hirota T., Takahashi A., Kubo M., Tsunoda T., Tomika K., Sakashita M. *et al.* Genome-wide
446 association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese
447 population. *Nat Genet* **44**, 1222-6 (2012).
448 11. Sun L.-D., Xiao F.L., Li Y., Zhou W.M., Tang H.Y., Tang X.F. *et al.* Genome-wide association
449 study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han
450 population. *Nature genetics* **43**, 690-4 (2011).
451 12. Esparza-Gordillo J., Weidinger S., Folster-Holst R., Bauerfeind A., Ruschendorf F., Patone G.
452 *et al.* A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nature*
453 *genetics* **41**, 596-601 (2009).

13. Paternoster L., Standl M. Waage J., Baurecht H., Hotze M., Strachan D.P. *et al*. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat Genet* **47**, 1449-56 (2015).
14. Boyd A., Golding J., Macleod J., Lawlor D.A., Fraser A., Henderson J. *et al*. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
15. Purvis D.J., Thompson J.M., Clark P.M., Robinson E., Black P.N., Wild C.J. *et al.* Risk factors for atopic dermatitis in New Zealand children at 3.5 years of age. *Br J Dermatol* **152**, 742-9 (2005).
16. Brown S.J., Relton C.L., Liao H., Zhao, Y., Sandilands A., Wilson I.J. *et al.* Filaggrin null mutations and childhood atopic eczema: a population-based case-control study. *J Allergy Clin Immunol* **121**, 940-46 e3 (2008).
17. Wijga A.H., Kerkhof M., Gehring U., de Jongste J.C., Postma D.S., Aalberse R.C. *et al.* Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. *Int J Epidemiol* **43**, 527-35 (2014).
18. Munafo M.R., Heron J. & Araya R. Smoking patterns during pregnancy and postnatal period and depressive symptoms. *Nicotine Tob Res* **10**, 1609-20 (2008).
19. Granell R., Henderson A.J. & Sterne J.A. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth cohort. *J Allergy Clin Immunol* **138**, 1060-1070 e11 (2016).
20. Muthén L.K & Muthén B.O. *Mplus User's Guide. Seventh Edition.* (Muthén & Muthén, Los Angeles, CA, 1998-2012).
21. Heron J.C., Croudace T.J., Barker E.D., Tilling K. A comparison of approaches for assessing covariate effects in latent class analysis. *Longitudinal and Life Course Studies* 6, 420-434 (2015).
22. Belgrave D.C., Granell R., Simpson A., Guiver J., Bishop C., Buchan I. *et al*. Developmental Profiles of Eczema, Wheeze, and Rhinitis: Two Population-Based Birth Cohort Studies. PLoS Med 11:e1001748 (2014).
23. Flohr C., Johansson S.G., Wahlgren C.F. & Williams H. How atopic is atopic dermatitis? J Allergy Clin Immunol 114, 150-8 (2004).
24. Dirven-Meijer P.C., Glazenburg E.J., Mulder P.G. & Oranje A.P. Prevalence of atopic dermatitis in children younger than 4 years in a demarcated area in central Netherlands: the West Veluwe Study Group. Br J Dermatol 158, 846-7 (2008).
25. Govaere E., Van Gysel D., Verhamme K.M., Doli E., Oranje A.P. & De Baets F. The prevalence, characteristics of and risk factors for eczema in Belgian schoolchildren. Pediatr Dermatol 26, 129-38 (2009).
26. Moore M.M., Rifas-Shiman S.L., Rich-Edwards J.W., Kleinman K.P., Camargo C.A.Jr, Gold D.R. *et al*. Perinatal predictors of atopic dermatitis occurring in the first six months of life. Pediatrics 113, 468-74 (2004).
27. Wen H.J., Chen P.C., Chiang T.L., Lin S.J., Chuang Y.L. & Guo Y.L. Predicting risk for early infantile atopic dermatitis by hereditary and environmental factors. Br J Dermatol 161, 1166-72 (2009).
28. Ip S., Chung M., Raman G., Chew P., Magula N., DeVine D. *et al*. Breastfeeding and maternal and infant health outcomes in developed countries. Evid Rep Technol Assess (Full Rep), 1-186 (2007).
29. Bisgaard H., Simpson A., Palmer C.N., Bonnelykke K., McLean I., Mukhopadhyay S. *et al*. Gene-environment interaction in the onset of eczema in infancy: filaggrin loss-of-function mutations enhanced by neonatal cat exposure. PLoS Med 5, e131 (2008).
30. Schuttelaar M.L., Kerkhof M., Jonkman M.F., Koppelman G.H., Brunekreef B., de Jongste J.C. *et al*. Filaggrin mutations in the onset of eczema, sensitization, asthma, hay fever and the interaction with cat exposure. Allergy 64, 1758-65 (2009).

31. Henderson J., Northstone K., Lee S.P., Liao H., Zhao Y., Pembrey M. *et al*. The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study. J Allergy Clin Immunol 121, 872-7 e9 (2008).
32. Brown S.J., Sandilands A., Zhao Y., Liao H., Relton C.L., Meggitt S.J., *et al*. Prevalent and low-frequency null mutations in the filaggrin gene are associated with early-onset and persistent atopic eczema. J Invest Dermatol 128, 1591-4 (2008).
33. Koppelman G.H. & Gehring U. Data-driven Asthma Phenotypes in Childhood. Does the Environment Hold the Clue? Am J Respir Crit Care Med. 195(5):545-546 (2017)

Figure Legends

Figure 1. Prevalence and frequency of AD in UK and Dutch longitudinal cohorts. Plot of frequency (right hand axis) of AD cases (grey bars) and controls (white bars) and AD prevalence (black points & left hand axis) over 12 time points in ALSPAC and 10 time points in PIAMA. AD is defined by presence of typical rash.

Figure 2. Longitudinal classes identified using LLCA in two independent birth cohorts, (A) ALSPAC (n=9894), (B) PIAMA (n=3652).

Tables

Table I. Descriptions and prevalences of the classes in two independent cohorts.

Class	Description of class in ALSPAC	ALSPAC prevalence	PIAMA prevalence
Unaffected individuals or transient AD	64% of this class never had reported rash, others had one or two isolated occasions of rash; ~10% reported rash consistent with AD at 6-18 months and this declined with age.	58.0%	62.9%
Early-onset-persistent AD	At 30 months of age ~85% of this class had reported rash, increasing to >90% prevalence until 12 years; it then steadily declined to ~50% at 16.5 years.	7.3%	4.9%
Early-onset-late-resolving AD	In this class the prevalence of rash rose steeply to >95% at 30 months and then steadily declined to ~10% by 16.5 years.	7.0%	3.8%
Early-onset-early-resolving AD	~60% of children in this class had reported rash at 18 and 30 months, this declined to 10% by 6-7 years.	12.9%	15.4%
Mid-onset-resolving AD	In this class there was a 10-20% prevalence of rash until 30 months, steeply rising to 75% prevalence 5-6 years, and steadily declining to <10% prevalence by 16.5 years.	7.0%	6.5%
Late-onset-resolving AD	In this class approximately 30% reported rash at 18 months, declining to ~10% prevalence at 5-6 years, steadily rising to ~70% prevalence by 12 years and finally declining to 10% by 16.5 years.	7.9%	6.5%

Table II. Association results between risk factors and AD classes identified by LLCA.

Trait	Exposed / Total	Wald P	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
ALSPAC			**7.3%**	**7.0%**	**12.9%**	**7.0%**	**7.9%**
Female	4805/9875	**6E-17**	**1.56(1.29-1.89) p=6E-6**	0.97(0.77-1.22) p=0.811	**0.75(0.62-0.91) p=0.004**	**1.79(1.40-2.29) p=4E-6**	**1.90(1.48-2.44) p=4E-7**
Maternal eczema	3154/9722	**3E-43**	**3.16(2.60-3.83) p=4E-31**	**1.68(1.32-2.14) p=2E-5**	**2.00(1.65-2.44) p=4E-12**	**1.66(1.29-2.13) p=8E-5**	**1.75(1.36-2.25) p=1E-5**
Maternal asthma	1554/9721	**2E-4**	**1.54(1.22-1.95) p=3E-4**	**1.43(1.08-1.91) p=0.014**	1.23(0.95-1.58) p=0.112	1.10(0.79-1.53) p=0.566	1.02(0.73-1.44) p=0.891
Paternal asthma	245/1568	**0.030**	1.59(0.86-2.93) p=0.139	**2.53(1.30-4.91) p=0.006**	1.58(0.86-2.89) p=0.139	0.94(0.38-2.33) p=0.893	1.72(0.83-3.57) p=0.146
Breastfeeding	7019/9198	**9E-4**	**1.42(1.11-1.81) p=0.006**	**1.53(1.12-2.08) p=0.008**	1.22(0.97-1.54) p=0.093	1.04(0.78-1.37) p=0.803	1.04(0.78-1.38) p=0.800
Pet cat	2963/9511	0.179	0.88(0.71-1.09) p=0.226	1.14(0.89-1.45) p=0.291	0.92(0.74-1.13) p=0.427	0.81(0.61-1.06) p=0.125	1.26(0.98-1.62) p=0.073
PIAMA			**4.9%**	**3.8%**	**15.4%**	**6.5%**	**6.5%**
Female	1759/3652	**0.025**	1.06(0.75-1.49) p=0.743	0.63(0.40-1.00) p=0.051	0.89(0.64-1.24) p=0.494	0.94(0.65-1.37) p=0.753	**1.87(1.21-2.90) p=0.005**
Maternal asthma	259/3645	**0.001**	**1.94(1.11-3.40) p=0.021**	**3.14(1.76-5.61) p=1E-4**	1.33(0.70-2.51) p=0.385	0.96(0.41-2.24) p=0.932	1.38(0.65-2.93) p=0.406
Paternal asthma	272/3633	**0.002**	**2.69(1.66-4.36) p=6E-5**	0.91(0.34-2.46) p=0.854	1.19(0.63-2.25) p=0.585	1.72(0.94-3.14) p=0.076	1.17(0.53-2.61) p=0.697
Breastfeeding	2984/3614	0.888	1.00(0.63-1.56) p=0.983	1.34(0.70-2.57) p=0.377	1.19(0.75-1.89) p=0.461	0.97(0.60-1.56) p=0.886	0.88(0.52-1.48) p=0.634
Pet cat	1213/3651	0.151	0.73(0.50-1.06) p=0.098	0.80(0.50-1.29) p=0.367	0.72(0.50-1.04) p=0.081	1.00(0.68-1.47) p=0.996	0.67(0.42-1.07) p=0.094

'Wald P' is for the overall omnibus test. Individual p-values and effect sizes comparing each class with the 'unaffected/transient' class are also shown; results p<0.05 are shown in bold.

Table III. Association results between AD classes identified by LLCA and comorbidities.

Trait	Cases / Total	Wald P	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
ALSPAC			**7.3%**	**7.0%**	**15.4%**	**7.0%**	**7.9%**
asthma at 7 years of age	904/7859	**$2x10^{-50}$**	**5.50(4.28-7.05) $p=5x10^{-41}$**	**3.08(2.22-4.27) $p=2x10^{-11}$**	**1.56(1.09-2.24) p=0.015**	**2.23(1.53-3.26) $p=3x10^{-5}$**	**1.89(1.26-2.83) p=0.002**
asthma at 13 years of age	784/6752	**$7x10^{-58}$**	**7.19(5.48-9.42) $p=3x10^{-46}$**	**3.59(2.51-5.12) $p=2x10^{-12}$**	**1.79(1.20-2.65) p=0.004**	**3.41(2.35-4.96) $p=1x10^{-10}$**	**2.01(1.30-3.12) p=0.002**
Elevated IgE at 7 years of age	2057/4790	**$8x10^{-16}$**	**2.62(1.98-3.47) $p=1x10^{-11}$**	**1.92(1.38-2.68) $p=1x10^{-4}$**	1.15(0.88-1.51) p=0.310	**1.55(1.10-2.18) p=0.013**	1.38(0.99-1.93) p=0.059
PIAMA			**4.9%**	**3.8%**	**15.4%**	**6.5%**	**6.5%**
asthma at 7 years of age	94/3349	**$2x10^{-15}$**	**14.27(7.33-27.78) $p=5x10^{-15}$**	**5.92(2.31-15.16) $p=2x10^{-4}$**	**3.03(1.08-8.47) p=0.035**	0.60(0.03-13.62) p=0.750	1.73(0.38-7.88) p=0.478
asthma at 11 years of age	102/2639	**$7x10^{-11}$**	**15.35(6.86-34.35) $p=3x10^{-11}$**	**9.12(3.49-23.82) $p=6x10^{-6}$**	**4.91(1.66-14.55) p=0.004**	2.10(0.42-10.53) p=0.366	**5.70(1.98-16.43) p=0.001**
Elevated IgE at 8 years of age	723/1707	**$1x10^{-4}$**	**3.00(1.85-4.86) $p=8x10^{-6}$**	1.58(0.86-2.89) p=0.140	1.57(0.97-2.56) p=0.067	1.42(0.83-2.46) p=0.203	0.97(0.53-1.77) p=0.914

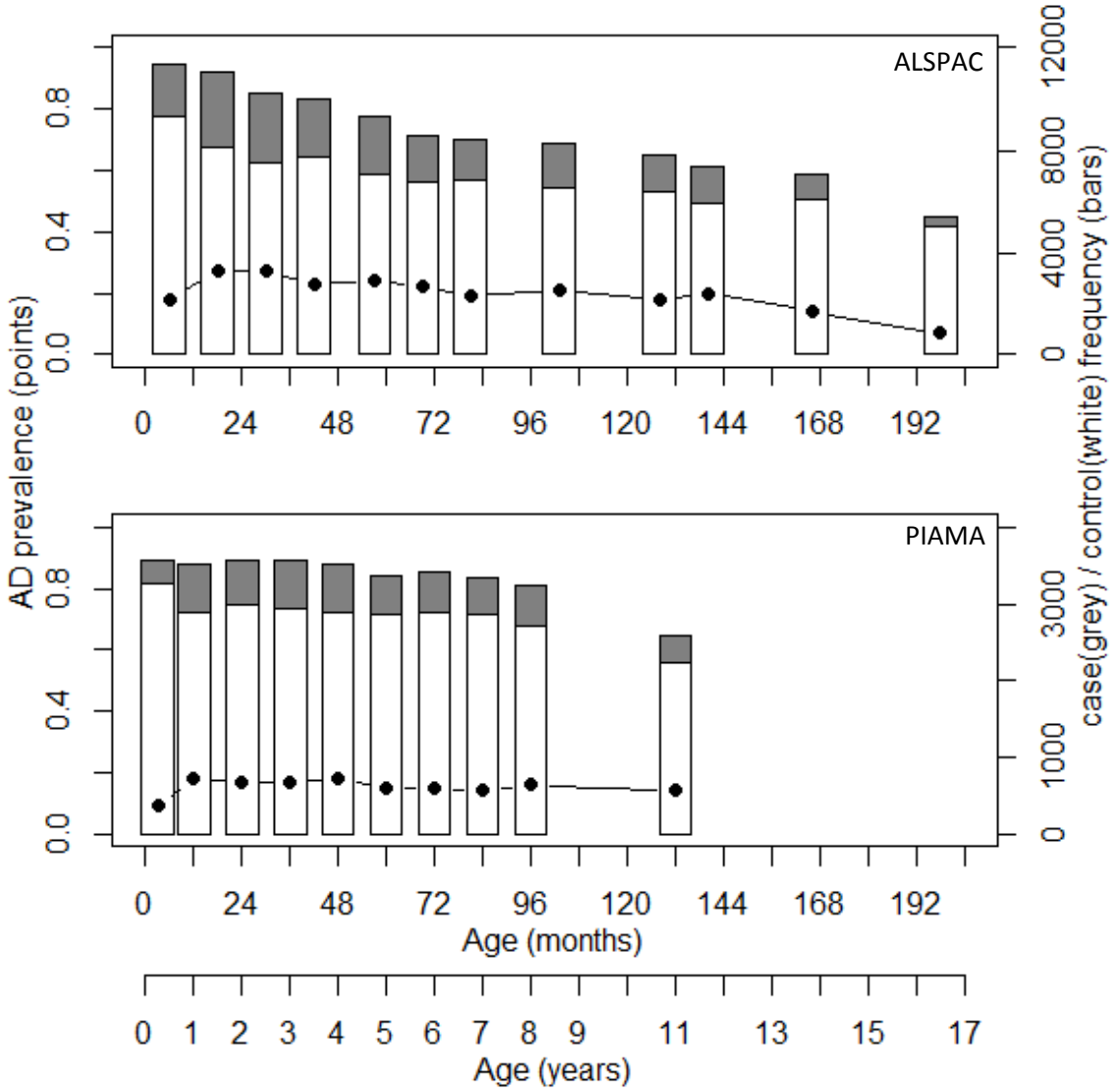
'Wald P' is for the overall omnibus test. Individual p-values and effect sizes comparing each class with the 'unaffected/transient' class are also shown; elevated IgE is defined as total IgE >75kU/L; results p<0.05 are shown in bold.

Table IV. Association results between genetic risk factors and AD classes identified by LLCA.

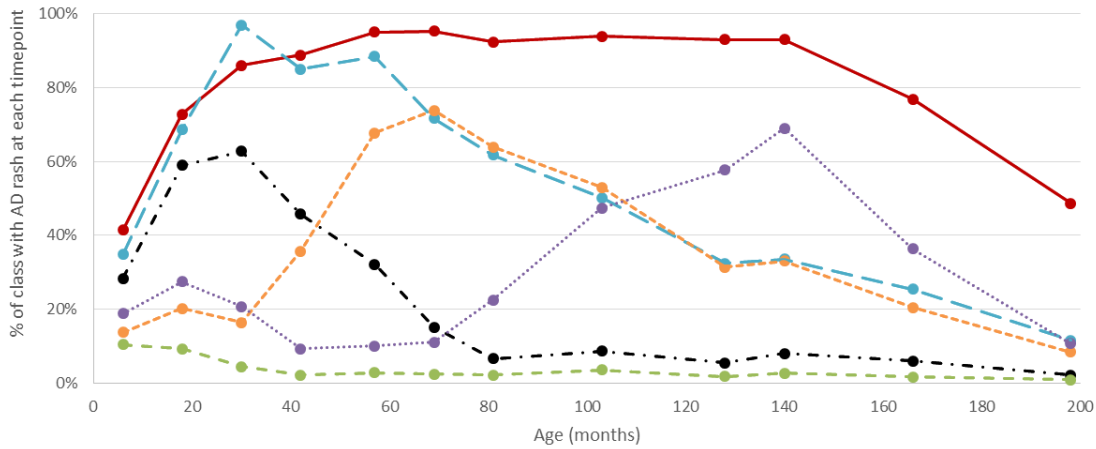
	# with risk genotype / total	Wald P	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
ALSPAC			**7.3%**	**7.0%**	**12.9%**	**7.0%**	**7.9%**
FLG null mutation	813/7570	**$4x10^{-28}$**	131/570 (23%)* **4.31(3.29-5.63)** **$p=2x10^{-26}$**	78/514 (15%)* **2.23(1.53-3.26)** **$p=3x10^{-5}$**	111/832 (13%)* **2.14(1.54-2.98)** **$p=7x10^{-6}$**	54/467 (12%)* 1.48(0.92-2.39) p=0.109	70/499 (14%)* **2.30(1.57-3.38)** **$p=2x10^{-5}$**
Genetic risk score (all other variants)	total N=6497	**$8x10^{-17}$**	**1.17(1.12-1.22)** **$p=2x10^{-13}$**	**1.08(1.04-1.13)** **$p=4x10^{-4}$**	1.02(0.99-1.06) p=0.222	**1.06(1.00-1.12)** **p=0.042**	1.01(0.96-1.06) p=0.758
PIAMA			**4.9%**	**3.8%**	**15.4%**	**6.5%**	**6.5%**
FLG null mutation	117/1516	**$6x10^{-4}$**	7/74 (10%)* 1.34(0.49-3.67) p=0.563	14/60 (23%)* **5.63(2.65-11.95)** **$p=7x10^{-6}$**	11/159 (7%)* 0.87(0.25-3.03) p=0.821	7/96 (7%)* 0.95(0.26-3.45) p=0.942	10/95 (11%)* 1.87(0.76-4.62) p=0.174
Genetic risk score (all other variants)	total N=1964	**$6x10^{-5}$**	**1.17(1.07-1.28)** **$p=5x10^{-4}$**	1.00(0.91-1.11) p=0.968	**1.16(1.08-1.25)** **$p=1x10^{-4}$**	**1.11(1.03-1.20)** **p=0.004**	1.08(0.98-1.18) p=0.111

'Wald P' is for the overall omnibus test. Individual p-values and effect sizes comparing each class with the 'unaffected/transient' class are also shown; results $p<0.05$ are shown in bold; Genetic risk score is defined by total number of risk alleles across the 23 AD-associated loci (other than *FLG*) identified by GWAS meta-analysis to date; OR represents the change in odds per risk allele for the genetic risk score or between carriers compared to non-carriers for the *FLG* mutations.

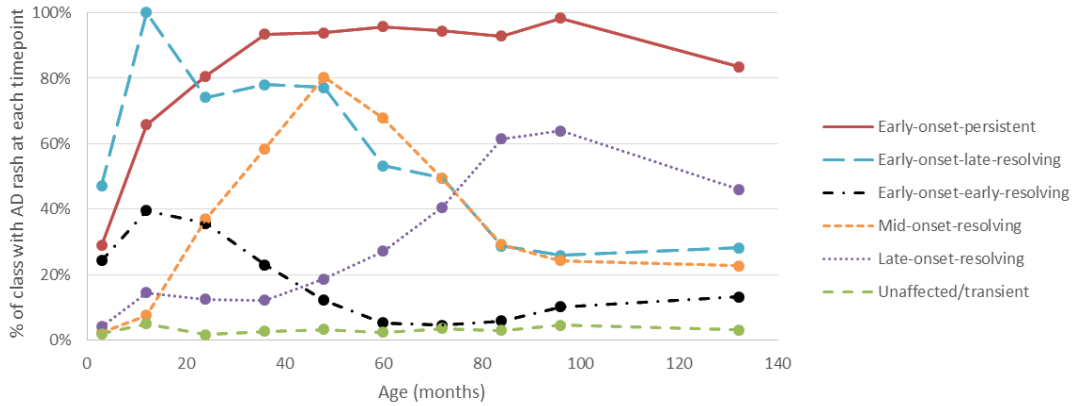
* In order to demonstrate approximate numbers of individuals with *FLG* null mutations in each class, individuals were assigned to most likely class. Given that the actual association analysis accounted for uncertainty in assignment of classes, these values are approximations for purposes of highlighting where power might be low. The approximate number with *FLG* null mutations / approximate total with *FLG* genotype data (%), within each class are shown. The 'unaffected/transient' groups had 8% (approx. 369/4712) and 7% (approx. 68/1032) with *FLG* mutations in ALSPAC and PIAMA, respectively.


ALSPAC
PIAMA
AD prevalence (points)
case(grey) / control(white) frequency (bars)
0.0
0.4
0.8
0
4000
8000
12000
0
1000
3000
0
24
48
72
96
120
144
168
192
Age (months)
0
1
2
3
4
5
6
7
8
9
11
13
15
17
Age (years)

A


100%
80%
60%
40%
20%
0%
% of class with AD rash at each timepoint
0
20
40
60
80
100
120
140
160
180
200
Age (months)

B


100%
80%
60%
40%
20%
0%
% of class with AD rash at each timepoint
0
20
40
60
80
100
120
140
Age (months)
Early-onset-persistent
Early-onset-late-resolving
Early-onset-early-resolving
Mid-onset-resolving
Late-onset-resolving
Unaffected/transient

Online Repository

The Avon Longitudinal Study of Parents and Children (ALSPAC)

This population-based birth cohort recruited 15,247 pregnant women resident in Avon, UK with expected dates of delivery from 1st April 1991 to 31st December 1992, resulting in 14,775 live births and 14,701 children who were alive at 1 year of age. Enrolment is described in detail in the cohort profile papers[1,2] and the study website contains details of all the data that are available through a fully searchable data dictionary[3]. Biological samples including DNA have been collected for 10,121 of the children from this cohort. Ethical approval for this study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees; written informed consent was provided by the parents/guardians.

DNA was obtained from blood samples using standard extraction techniques. The 4 most prevalent loss of function mutations in the gene encoding filaggrin (*FLG*) R501X, 2282del4, R2447X and S3247X were genotyped using KASP™ genotyping technology[4] (LGC Genomics, Hoddesdon, Hertfordshire, UK). Genome-wide genotyping was undertaken in 9912 children using the Illumina HumanHap550 quad SNP genotyping platform by 23andMe (Mountainview, California, USA), subcontracting to the Wellcome Trust Sanger Institute (Cambridge, UK) and the Laboratory Corporation of America (Burlington, NC, USA). Genotypes were imputed with Impute2 using 1000 Genomes phase 1 version 3 as a reference set as previously described[5].

Prevention and Incidence of Asthma and Mite Allergy birth cohort study (PIAMA)

The PIAMA study is a Dutch multicenter birth cohort that invited allergic and nonallergic pregnant women to participate; 4146 (53%) agreed and provided written informed consent (1327 allergic and 2819 nonallergic subjects). 183 were lost to follow-up before any data were obtained, so the study started with 3963 new-borns. Parents were sent International Study of Asthma and Allergies in Childhood–based questionnaires about their child's health, including asthma and eczema symptoms at 3, 12, 24, 36, 48, 60, 72, 84, 96 and 132 months after birth. All high-risk children and a sample of low-risk children were invited for clinical examinations at 4 years of age and at 8 years all children who were still in the study were invited. Full details of the study have been published previously[6]. The study protocol was approved by the medical ethics committees of the participating institutions and informed written parental consent was obtained for each participant.

Blood samples were collected from the children at age 4 and 8 years for DNA extraction. Children who did not participate in a clinical examination were invited to send a buccal swab by mail, and DNA was extracted from buccal swabs as described previously[7]. DNA was obtained from 2,162 children. Genome-wide genotyping was performed in three phases. The first phase was performed within the framework of the GABRIEL Consortium using an Illumina Human 610K quad array[8]. Genotypes were available from 172 children with asthma and from 187 controls after quality control. A second group of 268 children who were more extensively examined during follow up were genotyped with the Illumina HumanOmniExpress array. A final group of 1377 children was genotyped with the Illumina Human Omni Express Exome Array. SNPs were harmonized by base pair position annotated to genome build 37, name and annotation of strand for each platform. Discordant or duplicate SNPs or SNPs that showed large differences in allele frequencies (>15%) were removed. After quality control, a total of 1968 individuals remained and imputation was

performed per platform using IMPUTE 2.07 against the reference data set of the CEU panel of the 1000 Genomes project (version March 2012). SNPs of high quality (info-score IMPUTE ≥ 0.7) were merged into one dataset using GTOOL and used for further analysis. Dosages of imputed SNPs were predicted based on the following assumptions; 0 is the first homozygote, 0.5 is a heterozygote and 1 is the other homozygote. The info scores of the imputed SNPs were all ≥0.99 which provided reliable dosages estimates. From this dataset, 22 genotypes were available to calculate genetic risk scores for further analysis. Prevalent *FLG* loss-of-function mutations were genotyped at LGC Genomics as described above for the ALSPAC cohort.

LLCA – selection of best model

Longitudinal latent class analysis was carried out in Mplus. Models with between 2 and 7 classes were tested. Bayesian Information Criterion (BIC), entropy, and p-values from the Vuong-Lo-Mendell-Rubin (VLMR) likelihood ratio tests were compared to ascertain the models with the best fit.

ALSPAC – model fit

In the dataset of complete data (N=3480), the BIC (which is penalized for model complexity) was lowest for a 6 class model (Table E1 and Fig E1a) and the VLMR LR p-value showed evidence of improved fit for the 6-class model compared to a 5-class model (p=0.018). Entropy remained above 0.8 in all models, indicating good class delineation and individual class classifications were all >0.75, with <0.11 for all off-diagonals. The largest decrease in the BIC were seen from 2 to 4 classes (Fig E1a), with only small improvement in model fit between the 4- and 6-class models.

9894 individuals had data available from at least 6 of the 12 time points and the model fit parameters were broadly consistent with the smaller but more complete dataset (Table E1 and Fig E1b). The entropies for the 4- and 6-class models were 0.78 and 0.75, respectively indicating only small decreases in model fit compared to the smaller, but complete analysis and the individual classifications for each class, were all >0.7, with <0.13 for all off-diagonals. Comparison of results from the larger incomplete and smaller but complete datasets showed similar class assignment: only 3% of children changed best-fit class between the 6-class models in each analysis (Table E2) and the proportions of individuals in each class are very similar between models (Table E3)

Moving from the 6-class to the 4-class model, 99% of the persistent class (733/739) and the normal class (6111/6168) remained in these classes in the 4 class model. The majority of the early-onset-early-resolving (81%) and early-onset-late-resolving (61%) were in the 4-class early-onset class. The majority of the mid-onset (49%) and late-onset (87%) were in the 4-class late-onset class (Table E4).

PIAMA – model fit

In the dataset of complete data (N=2063), the BIC was lowest for a 4 class model (Table E5 & Fig E2a) and the VLMR LR p-value showed evidence of improved fit for the 4-class model compared to a 3-class model (p=0.015). However, there was only a small increase in BIC in the 6 class model (Table E5

& Fig E2a).

3652 individuals had data available from at least 5 of the 10 time points and the model fit parameters were broadly consistent with the smaller but more complete dataset (Table E5 and Fig E2b). The entropies for the 4- and 6-class models were 0.76 and 0.81, respectively indicating only small decreases in model fit compared to the smaller, but complete analysis and the proportions of individuals in each class are similar between the incomplete and complete analyses (Table E6).

Moving from the 6-class to the 4-class model, similar to ALSPAC, 99% of the persistent class (178/180) and the normal class (2468/2497) remained in these classes in the 4 class model. The majority of the early-onset-early-resolving (51%) and early-onset-late-resolving (68%) were in the 4-class early-onset class. The majority of the mid-onset (55%) and late-onset (91%) were in the 4-class late-onset class (Table E7).

Average latent class probability assignments

Average latent class probability assignments were >70% for all classes in the ALSPAC and PIAMA final models (Table E8). Unaffected and persistent classes had the highest probabilities (>80% in all cases). The largest probabilities for alternative class membership were between persistent and early-onset later-resolving in ALSPAC (12%) and early-onset early-resolving an unaffected in PIAMA (18%).

Bias –adjusted three step method for testing association of covariates with latent classes.

Association of risk factors and comorbidities with the latent classes were tested using a bias-adjusted three step analysis[9] to account for uncertainty in class assignment.
Following the LLCA (step 1), class assignment probabilities were exported to Stata version 14-MP[10] and logit constraints were derived to define the relationship between modal (or best) class (W) and latent classes (X) (which account for the uncertainty in modal class) (step 2). These logit constraints were then used in regression analyses in MPlus (step 3).

Supplementary References

1. Boyd A., Golding J., Macleod J., Lawlor D.A., Fraser A., Henderson J. *et al*. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 42, 111-27 (2013).

2. Fraser A., Macdonald-Wallis C., Tilling K., Boyd A., Golding J., Davey Smith G., *et al*. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol 42, 97-110 (2013).

3. http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/.

4. He, C., Holme, J. & Anthony, J. SNP genotyping: the KASP assay. Methods Mol Biol 1145, 75-86 (2014).

5. Paternoster L., Standl M. Waage J., Baurecht H., Hotze M., Strachan D.P. *et al*. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet 47, 1449-56 (2015).

6. Wijga A.H., Kerkhof M., Gehring U., de Jongste J.C., Postma D.S., Aalberse R.C. *et al*. Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. Int J Epidemiol 43, 527-35 (2014).

7. Bottema R.W., Reijmerink N.E., Kerkhof M., Koppelman G.H., Stelma F.F., Gerritsen J. *et al*. Interleukin 13, CD14, pet and tobacco smoke influence atopy in three Dutch cohorts: the allergenic study. Eur Respir J 32, 593-602 (2008).

8. Moffatt M.F., Gut I.G., Strachan D.P., Bouzigon E., Heath S., von Mutius E. *et al*. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363, 1211-21 (2010).

9. Heron J.C., Croudace T.J., Barker E.D., Tilling K. A comparison of approaches for assessing covariate effects in latent class analysis. Longitudinal and Life Course Studies 6, 420-434 (2015).

10. StataCorp. Stata Statistical Software: Release 14. (StataCorp LP., College Station, TX, 2015).

Figure Legends

Fig E1. Bayesian Information Criterion for ALSPAC model fit. (a) smaller dataset with complete data (N=3480), (b) larger dataset with ≥6 of the 12 time-points available (N=9894).

Fig E2. Bayesian Information Criterion across all PIAMA models, for (a) individuals with complete data (N=2063), and (b) individuals with incomplete data (N=3652)

Fig E3. 4-class model in ALSPAC

Fig E4. 4-class model in PIAMA

Tables

Table E1. LLCA model fit parameters in the ALSPAC cohort

	Individuals with complete data (N=3480)			Individuals with ≥6/12 time points available (N=9894)		
Number of classes	BIC	Entropy	VLMR LRT p-value	BIC	Entropy	VLMR LR p-value
2	33643	0.91	<0.0001	78475	0.88	<0.0001
3	32580	0.83	<0.0001	76525	0.79	<0.0001
4	**32182**	**0.83**	**<0.0001**	**75655**	**0.78**	**<0.0001**
5	32083	0.80	0.0394	75386	0.77	<0.0001
6	**32035**	**0.80**	**0.0177**	**75288**	**0.75**	**0.1289**
7	32049	0.80	0.0449	75212	0.75	0.0001

BIC, Bayesian Information Criterion; VLMR LRT, Vuong–Lo–Mendell–Rubin Likelihood ratio test. The 4- and 6-class models (**in bold**), both showed evidence of good model fit; the 6-class model had the lowest BIC value, but this was only a minor improvement in model fit compared to the 4-class model, according to VLMR LRT.

Table E2. Comparison of ALSPAC 6 class models constructed with individuals with complete data and constructed with individuals with incomplete data

incomplete / complete	P	EO-LR	EO-ER	MO-R	LO-R	unaffected
P	282	0	0	0	1	0
EO-LR	0	264	4	8	0	0
EO-ER	0	7	387	20	4	26
MO-R	5	0	0	193	3	1
LO-R	0	2	4	6	246	17
unaffected	0	0	0	6	2	1992

Incomplete data represents at least 6 of the 12 time-points available; the majority of individuals (diagonals, 97%) do not change classes; 116/3480 (3%) change classes.

In order to carry out this comparison individuals were assigned to their most likely class for the two models. Given that entropies are only 0.75-0.8, this is an approximation.

Table E3. Comparison of proportion of individuals in ALSPAC 6 class models constructed with complete data and constructed with individuals with incomplete data

Class	Incomplete analysis	Complete analysis
P	7.3%	7.9%
EO-LR	7.0%	8.1%
EO-ER	12.9%	14.5%
MO-R	7.0%	6.0%
LO-R	7.9%	8.7%
unaffected	58.0%	54.9%

Table E4. Comparison of ALSPAC 4- & 6-class models constructed with individuals with at least 6 of the 12 time-points available.

4 class 6 class	P	EO-R	LO-R	unaffected
P	733	2	4	0
EO-LR	246	401	11	0
EO-ER	0	876	2	206
MO-R	64	238	296	3
LO-R	12	23	545	45
unaffected	0	14	43	6111

In order to carry out this comparison individuals were assigned to their most likely class for the two models. Given that entropies are only 0.75-0.78, this is an approximation.

Table E5. LLCA model fit parameters in the PIAMA cohort

	Individuals with complete data (N=2063)			Individuals with ≥5/10 time points available (N=3652)		
Classes	BIC	Entropy	VLMR LR p-value	BIC	Entropy	VLMR LR p-value
2	14395	0.92	<0.0001	23850	0.91	<0.0001
3	14147	0.83	<0.0001	23476	0.82	<0.0001
4	**14067**	**0.84**	**0.0146**	**23327**	**0.81**	**0.0139**
5	14076	0.76	0.0888	23303	0.76	0.0675
6	**14105**	**0.73**	**0.0228**	**23312**	**0.76**	**0.0751**
7	14141	0.71	0.3658	23343	0.78	0.0057

Table E6. Comparison of proportion of individuals in PIAMA 6 class models constructed with complete data and constructed with individuals with incomplete data

Class	Incomplete analysis	Complete analysis
P	4.9%	4.8%
EO-LR	3.8%	4.5%
EO-ER	15.4%	12.4%
MO-R	6.5%	8.7%
LO-R	6.5%	11.8%
unaffected	62.9%	57.8%

Table E7. Comparison of PIAMA 4 & 6 class models constructed with individuals with at least 5 of the 10 time-points available.

4 class 6 class	P	EO-R	LO-R	unaffected
P	178	0	2	0
EO-LR	44	98	2	0
EO-ER	0	203	6	191
MO-R	26	76	126	0
LO-R	1	5	184	13
unaffected	0	0	29	2468

In order to carry out this comparison individuals were assigned to their most likely class for the two models. Given that entropies are only 0.76-0.81, this is an approximation.

Table E8. Average latent class probabilities for the most likely class membership (row) by latent class (column).

	ALSPAC						PIAMA					
	P	EO-LR	EO-ER	MO-R	LO-R	UA	P	EO-LR	EO-ER	MO-R	LO-R	UA
P	0.834	0.122	0.004	0.030	0.009	0.000	0.902	0.047	0.000	0.042	0.009	0.000
EO-LR	0.093	0.718	0.094	0.076	0.019	0.001	0.042	0.764	0.092	0.075	0.025	0.002
EO-ER	0.004	0.069	0.735	0.047	0.047	0.098	0.001	0.031	0.718	0.032	0.043	0.175
MO-R	0.042	0.048	0.070	0.716	0.081	0.043	0.03	0.022	0.081	0.739	0.084	0.044
LO-R	0.016	0.023	0.061	0.097	0.722	0.082	0.02	0.012	0.066	0.083	0.717	0.103
UA	0.000	0.001	0.054	0.012	0.034	0.898	0.000	0.000	0.092	0.008	0.021	0.879

UA=unaffected

Table E9. Association results for all four LCA models in ALSPAC (incomplete/complete & 4-/6-class models)

INCOMPLETE CASE 6 CLASS

	N	omni_p	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
			7.3%	7.0%	12.9%	7.0%	7.9%
female	9875	**6E-17**	**1.56(1.29-1.89) p=6E-6**	0.97(0.77-1.22) p=0.811	**0.75(0.62-0.91) p=0.004**	**1.79(1.40-2.29) p=4E-6**	**1.90(1.48-2.44) p=4E-7**
maternal eczema	9722	**3E-43**	**3.16(2.60-3.83) p=4E-31**	**1.68(1.32-2.14) p=2E-5**	**2.00(1.65-2.44) p=4E-12**	**1.66(1.29-2.13) p=8E-5**	**1.75(1.36-2.25) p=1E-5**
maternal asthma	9721	**2E-4**	**1.54(1.22-1.95) p=3E-4**	**1.43(1.08-1.91) p=0.014**	1.23(0.95-1.58) p=0.112	1.10(0.79-1.53) p=0.566	1.02(0.73-1.44) p=0.891
paternal asthma	1568	**0.030**	1.59(0.86-2.93) p=0.139	**2.53(1.30-4.91) p=0.006**	1.58(0.86-2.89) p=0.139	0.94(0.38-2.33) p=0.893	1.72(0.83-3.57) p=0.146
breastfeeding	9198	**9E-4**	**1.42(1.11-1.81) p=0.006**	**1.53(1.12-2.08) p=0.008**	1.22(0.97-1.54) p=0.093	1.04(0.78-1.37) p=0.803	1.04(0.78-1.38) p=0.800
cat	9511	0.179	0.88(0.71-1.09) p=0.226	1.14(0.89-1.45) p=0.291	0.92(0.74-1.13) p=0.427	0.81(0.61-1.06) p=0.125	1.26(0.98-1.62) p=0.073
FLG	7570	**4E-28**	**4.31(3.29-5.63) p=2E-26**	**2.23(1.53-3.26) p=3E-5**	**2.14(1.54-2.98) p=7E-6**	1.48(0.92-2.39) p=0.109	**2.30(1.57-3.38) p=2E-5**
genetic risk score	6497	**8E-17**	**1.17(1.12-1.22) p=2E-13**	**1.08(1.04-1.13) p=4E-4**	1.02(0.99-1.06) p=0.222	**1.06(1.00-1.12) p=0.042**	1.01(0.96-1.06) p=0.758
asthma at age 7	7859	**2E-50**	**5.50(4.28-7.05) p=5E-41**	**3.08(2.22-4.27) p=2E-11**	**1.56(1.09-2.24) p=0.015**	**2.23(1.53-3.26) p=3E-5**	**1.89(1.26-2.83) p=0.002**
asthma at age 13	6752	**7E-58**	**7.19(5.48-9.42) p=3E-46**	**3.59(2.51-5.12) p=2E-12**	**1.79(1.20-2.65) p=0.004**	**3.41(2.35-4.96) p=1E-10**	**2.01(1.30-3.12) p=0.002**
IgE>75 at age 7	4790	**8E-16**	**2.62(1.98-3.47) p=1E-11**	**1.92(1.38-2.68) p=1E-4**	1.15(0.88-1.51) p=0.310	**1.55(1.10-2.18) p=0.013**	1.38(0.99-1.93) p=0.059

INCOMPLETE CASE 4 CLASS

omni_p	Early-onset persistent	Early-onset resolving	Late-onset resolving
	10.7%	16.5%	10.9%
2E-12	**1.44(1.24-1.67) p=1E-6**	0.94(0.81-1.09) p=0.393	**1.96(1.60-2.39) p=4E-11**
3E-41	**2.46(2.12-2.85) p=2E-32**	**1.87(1.60-2.17) p=5E-16**	**1.65(1.35-2.02) p=1E-6**
5E-4	**1.41(1.17-1.70) p=4E-4**	**1.38(1.14-1.66) p=8E-4**	1.02(0.78-1.35) p=0.864
0.086	**1.88(1.23-2.87) p=0.004**	1.43(0.88-2.32) p=0.150	1.24(0.64-2.39) p=0.524
4E-4	**1.46(1.20-1.77) p=1E-4**	**1.31(1.09-1.58) p=0.003**	0.96(0.77-1.20) p=0.730
0.693	0.91(0.78-1.07) p=0.268	0.91(0.78-1.07) p=0.278	1.08(0.88-1.33) p=0.436
6E-28	**3.50(2.81-4.35) p=2E-29**	**2.03(1.57-2.62) p=7E-8**	**1.83(1.31-2.55) p=4E-4**
6E-16	**1.14(1.11-1.18) p=2E-19**	**1.03(1.00-1.06) p=0.033**	1.03(0.99-1.07) p=0.153
2E-49	**4.83(3.94-5.91) p=1E-52**	**1.96(1.52-2.51) p=2E-7**	**1.76(1.27-2.45) p=8E-4**
9E-59	**6.41(5.16-7.97) p=7E-63**	**2.44(1.87-3.20) p=9E-11**	**2.23(1.57-3.15) p=6E-6**
4E-15	**2.59(2.08-3.23) p=2E-17**	**1.30(1.06-1.61) p=0.012**	1.12(0.85-1.47) p=0.417

COMPLETE CASE 6 CLASS

	N	omni_p	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
			7.9%	8.1%	14.5%	6.0%	8.7%
female	3475	**7E-7**	**1.39(1.05-1.84) p=0.022**	1.04(0.77-1.42) p=0.781	**0.66(0.49-0.89) p=0.006**	**1.63(1.10-2.41) p=0.014**	**2.00(1.41-2.83) p=9E-5**
maternal eczema	3473	**3E-20**	**3.35(2.51-4.47) p=2E-16**	**1.89(1.37-2.58) p=8E-5**	**2.14(1.60-2.86) p=3E-7**	**1.97(1.35-2.88) p=5E-4**	1.26(0.89-1.78) p=0.194
maternal asthma	3473	**0.006**	**1.77(1.26-2.49) p=0.001**	**1.53(1.04-2.25) p=0.031**	1.06(0.71-1.58) p=0.781	1.31(0.80-2.14) p=0.282	1.09(0.69-1.71) p=0.713
paternal asthma	964	**0.042**	2.01(0.99-4.06) p=0.052	**2.46(1.28-4.76) p=0.007**	1.60(0.79-3.25) p=0.193	1.19(0.38-3.73) p=0.767	2.17(0.92-5.14) p=0.077
breastfeeding	3435	**5E-4**	1.13(0.78-1.63) p=0.514	**1.91(1.15-3.17) p=0.012**	**1.75(1.11-2.75) p=0.015**	1.45(0.84-2.52) p=0.186	**0.64(0.44-0.92) p=0.015**
cat	3434	0.126	0.87(0.64-1.18) p=0.373	0.89(0.64-1.24) p=0.488	**0.64(0.46-0.90) p=0.010**	0.76(0.50-1.16) p=0.211	1.00(0.71-1.40) p=0.977
FLG	2994	**6E-11**	**4.14(2.75-6.22) p=9E-12**	**2.57(1.55-4.24) p=2E-4**	**2.60(1.63-4.15) p=7E-5**	1.31(0.57-3.02) p=0.521	**2.29(1.35-3.89) p=0.002**
genetic risk score	2655	**2E-7**	**1.18(1.11-1.25) p=6E-8**	1.05(0.99-1.11) p=0.113	1.03(0.98-1.09) p=0.218	1.08(0.99-1.18) p=0.076	0.96(0.90-1.02) p=0.182
asthma at age 7	3413	**2E-25**	**5.39(3.82-7.61) p=1E-21**	**3.35(2.23-5.02) p=5E-9**	1.24(0.73-2.11) p=0.430	**2.16(1.24-3.76) p=0.007**	1.15(0.61-2.16) p=0.670
asthma at age 13	3455	**1E-35**	**6.65(4.73-9.33) p=7E-28**	**4.24(2.88-6.24) p=2E-13**	1.33(0.79-2.25) p=0.289	**3.43(2.12-5.55) p=5E-7**	1.44(0.81-2.56) p=0.217
IgE>75 at age 7	2130	**1E-8**	**2.67(1.84-3.87) p=2E-7**	**2.12(1.42-3.19) p=3E-4**	1.11(0.77-1.60) p=0.574	**1.68(1.04-2.70) p=0.033**	1.19(0.78-1.82) p=0.408

COMPLETE CASE 4 CLASS

omni_p	Early-onset persistent	Early-onset resolving	Late-onset resolving
	11.4%	17.1%	11.4%
5E-4	**1.35(1.07-1.70) p=0.010**	0.93(0.74-1.17) p=0.540	**1.82(1.37-2.43) p=4E-5**
2E-21	**2.86(2.27-3.61) p=1E-18**	**2.14(1.70-2.69) p=9E-11**	**1.51(1.13-2.01) p=0.005**
0.003	**1.81(1.37-2.39) p=3E-5**	1.26(0.93-1.71) p=0.131	1.17(0.80-1.70) p=0.430
0.018	**2.50(1.49-4.20) p=5E-4**	1.58(0.89-2.80) p=0.118	1.94(0.93-4.06) p=0.079
0.008	1.17(0.86-1.59) p=0.327	**1.93(1.33-2.82) p=6E-4**	0.75(0.54-1.04) p=0.089
0.116	0.90(0.70-1.15) p=0.408	**0.68(0.52-0.88) p=0.004**	1.04(0.78-1.40) p=0.773
4E-11	**3.48(2.49-4.88) p=4E-13**	**2.22(1.53-3.23) p=3E-5**	1.52(0.91-2.52) p=0.110
2E-6	**1.14(1.09-1.20) p=2E-8**	**1.05(1.01-1.10) p=0.026**	1.00(0.95-1.05) p=0.970
4E-25	**5.13(3.83-6.87) p=5E-28**	**1.95(1.36-2.79) p=3E-4**	1.11(0.64-1.93) p=0.704
3E-37	**6.99(5.24-9.33) p=6E-40**	**2.80(2.00-3.91) p=2E-9**	1.54(0.94-2.54) p=0.087
1E-8	**2.70(1.98-3.68) p=3E-10**	**1.61(1.20-2.15) p=0.001**	1.02(0.71-1.46) p=0.913

p<0.05 shown in bold

Table E10. Association results for 4- and 6- class models in PIAMA

PIAMA 6 CLASS

var	N	omni_p	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
			4.9%	**3.8%**	**15.4%**	**6.5%**	**6.5%**
female	3652	0.025	1.06(0.75-1.49) p=0.743	0.63(0.40-1.00) p=0.051	0.89(0.64-1.24) p=0.494	0.94(0.65-1.37) p=0.753	**1.87(1.21-2.90) p=0.005**
maternal asthma	3645	0.001	**1.94(1.11-3.40) p=0.021**	**3.14(1.76-5.61) p=1E-4**	1.33(0.70-2.51) p=0.385	0.96(0.41-2.24) p=0.932	1.38(0.65-2.93) p=0.406
paternal asthma	3633	0.002	**2.69(1.66-4.36) p=6E-5**	0.91(0.34-2.46) p=0.854	1.19(0.63-2.25) p=0.585	1.72(0.94-3.14) p=0.076	1.17(0.53-2.61) p=0.697
breastfeeding	3614	0.888	1.00(0.63-1.56) p=0.983	1.34(0.70-2.57) p=0.377	1.19(0.75-1.89) p=0.461	0.97(0.60-1.56) p=0.886	0.88(0.52-1.48) p=0.634
cat	3651	0.151	0.73(0.50-1.06) p=0.098	0.80(0.50-1.29) p=0.367	0.72(0.50-1.04) p=0.081	1.00(0.68-1.47) p=0.996	0.67(0.42-1.07) p=0.094
FLG	1516	6E-4	1.34(0.49-3.67) p=0.563	**5.63(2.65-11.95) p=7E-6**	0.87(0.25-3.03) p=0.821	0.95(0.26-3.45) p=0.942	1.87(0.76-4.62) p=0.174
genetic risk score	1964	6E-5	**1.17(1.07-1.28) p=5E-4**	1.00(0.91-1.11) p=0.968	**1.16(1.08-1.25) p=1E-4**	**1.11(1.03-1.20) p=0.004**	1.08(0.98-1.18) p=0.111
asthma at age 7	3349	2E-15	**14.27(7.33-27.78) p=5E-15**	**5.92(2.31-15.16) p=2E-4**	**3.03(1.08-8.47) p=0.035**	0.60(0.03-13.62) p=0.750	1.73(0.38-7.88) p=0.478
asthma at age 11	2639	7E-11	**15.35(6.86-34.35) p=3E-11**	**9.12(3.49-23.82) p=6E-6**	**4.91(1.66-14.55) p=0.004**	2.10(0.42-10.53) p=0.366	**5.70(1.98-16.43) p=0.001**
IgE>75 at age 8	1707	1E-4	**3.00(1.85-4.86) p=8E-6**	1.58(0.86-2.89) p=0.140	1.57(0.97-2.56) p=0.067	1.42(0.83-2.46) p=0.203	0.97(0.53-1.77) p=0.914

PIAMA 4 CLASS

var		omni_p	Early-onset persistent	Early-onset resolving	Late-onset resolving
			6.6%	**11.4%**	**10.2%**
female	3652	0.004	1.00(0.75-1.34) p=0.984	**0.64(0.47-0.87) p=0.004**	**1.72(1.24-2.39) p=0.001**
maternal asthma	3645	0.007	**2.07(1.29-3.31) p=0.002**	**1.91(1.18-3.07) p=0.008**	1.51(0.85-2.69) p=0.158
paternal asthma	3633	0.010	**2.30(1.51-3.51) p=1E-4**	1.12(0.64-1.94) p=0.696	1.05(0.56-1.96) p=0.875
breastfeeding	3614	1.000	1.01(0.68-1.50) p=0.952	1.04(0.70-1.53) p=0.860	0.95(0.63-1.43) p=0.807
cat	3651	0.441	0.86(0.63-1.17) p=0.336	0.75(0.54-1.03) p=0.077	0.87(0.62-1.23) p=0.435
FLG	1516	0.192	1.99(0.97-4.09) p=0.062	**2.15(1.04-4.42) p=0.038**	1.36(0.60-3.12) p=0.464
genetic risk score	1964	0.005	**1.13(1.05-1.21) p=0.001**	**1.08(1.01-1.15) p=0.019**	1.05(0.98-1.12) p=0.173
asthma at age 7	3349	1E-12	**8.77(5.13-14.97) p=2E-15**	**3.45(1.70-7.00) p=6E-4**	0.46(0.03-7.05) p=0.579
asthma at age 11	2639	3E-10	**8.43(4.61-15.42) p=4E-12**	**5.47(2.76-10.85) p=1E-6**	**2.80(1.12-7.03) p=0.028**
IgE>75 at age 8	1707	3E-6	**2.64(1.74-4.00) p=5E-6**	**2.30(1.49-3.55) p=2E-4**	1.04(0.66-1.65) p=0.854

p<0.05 shown in bold

Table E11. Association results for individual SNPs with 4- and 6- class models in ALSPAC & PIAMA

ALSPAC 6 CLASS

var	Chr	Nearest gene	omni_p	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
				7.3%	7.0%	12.9%	7.0%	7.9%
rs10214237_t	5	*IL7R*	0.852	1.01(0.83-1.22) p=0.960	1.09(0.87-1.37) p=0.434	1.00(0.82-1.21) p=0.984	1.08(0.86-1.37) p=0.496	1.11(0.88-1.39) p=0.379
rs1057258_c	2	*INPP5D*	**0.011**	1.19(0.93-1.52) p=0.162	0.87(0.68-1.11) p=0.256	**0.80(0.65-0.99) p=0.039**	0.92(0.70-1.22) p=0.577	**0.73(0.57-0.93) p=0.011**
rs10995251_c	10	*ZNF365*	0.711	0.96(0.82-1.14) p=0.664	1.19(0.96-1.47) p=0.107	0.95(0.81-1.12) p=0.571	1.01(0.81-1.26) p=0.908	0.99(0.81-1.20) p=0.896
rs112111458_a	2	*CD207*	0.581	0.97(0.77-1.22) p=0.799	1.31(0.95-1.81) p=0.096	0.95(0.75-1.20) p=0.681	0.89(0.67-1.17) p=0.387	0.94(0.71-1.24) p=0.665
rs11657987_t	17	*PGS1*	0.613	1.11(0.94-1.33) p=0.225	1.08(0.88-1.32) p=0.461	1.04(0.89-1.23) p=0.596	1.10(0.88-1.37) p=0.396	1.03(0.86-1.24) p=0.742
rs12153855_t	6	*TNXB*	0.059	**1.35(1.01-1.80) p=0.043**	1.16(0.84-1.61) p=0.369	1.02(0.77-1.33) p=0.908	**1.55(1.02-2.36) p=0.040**	1.13(0.81-1.57) p=0.468
rs12295535_t	11	*PRR5L*	**0.026**	**1.54(1.02-2.34) p=0.040**	1.55(0.94-2.54) p=0.086	1.19(0.74-1.93) p=0.468	1.19(0.68-2.06) p=0.543	0.39(0.13-1.17) p=0.094
rs16948048_g	17	*ZNF652*	0.214	0.97(0.82-1.15) p=0.762	**1.21(1.00-1.47) p=0.046**	0.93(0.79-1.10) p=0.404	0.91(0.73-1.12) p=0.355	0.88(0.71-1.10) p=0.263
rs17389644_a	4	*IL21*	0.602	1.07(0.88-1.32) p=0.495	1.18(0.94-1.49) p=0.153	1.08(0.89-1.32) p=0.444	1.03(0.80-1.33) p=0.794	1.00(0.78-1.29) p=0.981
rs17881320_t	17	*STAT3*	**0.002**	**1.67(1.28-2.18) p=1E-4**	0.95(0.62-1.44) p=0.801	1.02(0.74-1.39) p=0.905	1.39(0.97-2.00) p=0.076	1.05(0.73-1.50) p=0.792
rs2041733_t	16	*CLEC16A*	0.558	1.03(0.88-1.22) p=0.703	1.03(0.85-1.25) p=0.785	1.06(0.90-1.26) p=0.493	1.19(0.97-1.46) p=0.099	0.92(0.75-1.13) p=0.426
rs2143950_t	14	*PPP2R3C*	**0.003**	**1.32(1.07-1.62) p=0.008**	1.23(0.96-1.57) p=0.099	0.86(0.67-1.10) p=0.219	1.28(0.99-1.65) p=0.063	1.13(0.88-1.45) p=0.344
rs2164983_a	19	*ACTL9*	**6E-5**	**1.55(1.28-1.89) p=1E-5**	1.18(0.90-1.54) p=0.228	**1.27(1.01-1.58) p=0.037**	1.29(0.98-1.69) p=0.067	1.02(0.75-1.40) p=0.901
rs2227483_t	12	*IL22*	0.593	1.12(0.94-1.33) p=0.207	0.88(0.71-1.08) p=0.217	1.03(0.87-1.22) p=0.698	1.07(0.86-1.32) p=0.555	1.10(0.90-1.35) p=0.361
rs2228145_c	1	*IL6R*	**0.002**	**1.36(1.15-1.60) p=4E-4**	1.07(0.88-1.30) p=0.520	0.95(0.80-1.14) p=0.594	1.05(0.85-1.28) p=0.658	0.87(0.71-1.07) p=0.193
rs2897442_c	5	*KIF3A*	**0.007**	**1.35(1.13-1.62) p=9E-4**	0.91(0.72-1.15) p=0.443	0.95(0.78-1.16) p=0.621	1.18(0.95-1.47) p=0.131	0.88(0.70-1.11) p=0.286
rs479844_g	11	*OVOL1*	**0.002**	**1.29(1.08-1.53) p=0.004**	**1.30(1.07-1.59) p=0.008**	1.07(0.90-1.28) p=0.419	0.98(0.76-1.24) p=0.840	1.10(0.90-1.33) p=0.363
rs6010620_g	20	*RTEL1*	0.174	**1.24(1.01-1.51) p=0.038**	1.18(0.92-1.51) p=0.200	1.07(0.89-1.29) p=0.468	1.02(0.81-1.29) p=0.840	1.04(0.83-1.31) p=0.721
rs6473227_c	8	*ZBTB10*	0.208	1.10(0.93-1.30) p=0.259	1.10(0.90-1.35) p=0.361	1.13(0.95-1.34) p=0.159	1.18(0.95-1.46) p=0.142	1.12(0.91-1.37) p=0.299
rs6602364_g	10	*IL2RA*	0.935	0.95(0.81-1.12) p=0.552	1.02(0.83-1.26) p=0.834	0.99(0.83-1.18) p=0.928	0.89(0.71-1.13) p=0.351	0.98(0.79-1.21) p=0.854
rs7127307_t	11	*ETS1*	0.465	1.16(0.98-1.37) p=0.078	0.97(0.81-1.17) p=0.752	1.04(0.88-1.22) p=0.673	1.02(0.84-1.23) p=0.864	1.15(0.94-1.40) p=0.165
rs7146581_c	14	*TRAF3*	0.215	1.22(0.99-1.50) p=0.068	1.04(0.82-1.31) p=0.758	1.22(0.99-1.50) p=0.064	1.06(0.83-1.34) p=0.655	1.14(0.90-1.44) p=0.282
rs7927894_t	11	*C11orf30*	**0.023**	**1.27(1.08-1.50) p=0.005**	1.00(0.82-1.22) p=0.976	0.91(0.77-1.07) p=0.253	1.10(0.89-1.35) p=0.370	1.13(0.92-1.38) p=0.232

PIAMA 6 CLASS

var	Chr	Nearest gene	omni_p	Early-onset persistent	Early-onset late-resolving	Early-onset early-resolving	Mid-onset resolving	Late-onset resolving
				4.9%	3.8%	15.4%	6.5%	6.5%
rs10214237_t	5	*IL7R*	0.486	1.07(0.74-1.55) p=0.724	0.78(0.53-1.15) p=0.211	0.96(0.67-1.40) p=0.845	1.22(0.85-1.75) p=0.274	1.25(0.83-1.88) p=0.279
rs1057258_c	2	*INPP5D*	0.405	0.83(0.53-1.28) p=0.391	0.77(0.50-1.18) p=0.234	1.00(0.65-1.53) p=1.000	0.78(0.52-1.18) p=0.243	0.74(0.45-1.20) p=0.218
rs10995251_c	10	*ZNF365*	0.480	1.22(0.85-1.76) p=0.280	0.94(0.62-1.42) p=0.779	1.35(0.95-1.93) p=0.096	0.98(0.68-1.40) p=0.913	0.83(0.53-1.29) p=0.401
rs112111458_a	2	*CD207*	0.454	1.42(0.83-2.42) p=0.202	0.68(0.41-1.15) p=0.155	0.99(0.59-1.68) p=0.979	1.36(0.72-2.57) p=0.342	0.85(0.51-1.45) p=0.558
rs11657987_t	17	*PGS1*	0.421	1.11(0.82-1.48) p=0.505	0.71(0.50-1.02) p=0.065	1.14(0.82-1.58) p=0.433	1.13(0.82-1.56) p=0.450	1.04(0.72-1.49) p=0.851
rs12153855_t	6	*TNXB*	0.742	1.34(0.67-2.69) p=0.405	1.15(0.52-2.57) p=0.733	1.35(0.74-2.46) p=0.333	1.20(0.61-2.38) p=0.596	1.60(0.59-4.34) p=0.359
rs12295535_t	11	*PRR5L*	0.694	0.25(0.03-2.40) p=0.228	1.37(0.50-3.74) p=0.544	1.15(0.44-3.01) p=0.779	1.50(0.66-3.43) p=0.337	1.44(0.53-3.88) p=0.471
rs16948048_g	17	*ZNF652*	0.071	1.17(0.84-1.62) p=0.350	1.17(0.76-1.79) p=0.482	0.95(0.68-1.31) p=0.746	**1.58(1.15-2.16) p=0.005**	0.96(0.65-1.42) p=0.834
rs17389644_a	4	*IL21*	0.141	1.43(0.98-2.08) p=0.062	1.44(0.93-2.23) p=0.098	0.91(0.63-1.30) p=0.595	1.16(0.73-1.85) p=0.537	1.04(0.61-1.77) p=0.895
rs17881320_t	17	*STAT3*	0.078	**1.81(1.08-3.06) p=0.026**	1.04(0.44-2.47) p=0.933	**1.92(1.16-3.16) p=0.011**	1.27(0.69-2.35) p=0.448	0.75(0.27-2.09) p=0.584
rs2041733_t	16	*CLEC16A*	0.094	1.38(0.98-1.94) p=0.065	0.96(0.66-1.38) p=0.809	1.15(0.83-1.59) p=0.390	**1.46(1.07-2.01) p=0.018**	1.15(0.78-1.69) p=0.490
rs2143950_t	14	*PPP2R3C*	0.297	1.37(0.92-2.02) p=0.118	0.99(0.58-1.68) p=0.956	1.20(0.80-1.81) p=0.384	0.63(0.36-1.09) p=0.098	0.90(0.52-1.58) p=0.721
rs2164983_a	19	*ACTL9*	0.100	1.30(0.83-2.03) p=0.247	1.49(0.92-2.40) p=0.102	0.76(0.47-1.22) p=0.255	0.92(0.56-1.51) p=0.739	1.55(0.96-2.50) p=0.074
rs2227483_t	12	*IL22*	**0.008**	**1.36(1.01-1.82) p=0.043**	0.93(0.62-1.40) p=0.735	**1.63(1.15-2.31) p=0.006**	1.00(0.70-1.44) p=0.996	**1.89(1.19-2.99) p=0.007**
rs2228145_c	1	*IL6R*	0.368	0.97(0.70-1.34) p=0.833	1.27(0.89-1.82) p=0.184	1.30(0.96-1.77) p=0.094	0.88(0.63-1.24) p=0.480	1.14(0.80-1.64) p=0.473
rs2897442_c	5	*KIF3A*	0.542	1.04(0.71-1.52) p=0.832	1.23(0.81-1.85) p=0.327	0.73(0.47-1.14) p=0.167	1.03(0.70-1.51) p=0.894	1.19(0.80-1.77) p=0.394
rs479844_g	11	*OVOL1*	0.341	**1.47(1.02-2.10) p=0.038**	0.99(0.68-1.45) p=0.971	1.08(0.78-1.49) p=0.643	1.15(0.82-1.60) p=0.417	0.86(0.58-1.28) p=0.451
rs6010620_g *	20	*RTEL1*	0.359	**1.63(1.06-2.52) p=0.027**	0.94(0.61-1.44) p=0.768	1.20(0.86-1.68) p=0.287	1.00(0.69-1.45) p=0.992	1.00(0.65-1.54) p=0.986
rs6602364_g	10	*IL2RA*	**0.033**	0.98(0.71-1.34) p=0.892	0.71(0.45-1.12) p=0.143	1.22(0.86-1.72) p=0.261	1.03(0.71-1.49) p=0.891	**1.60(1.17-2.20) p=0.003**
rs7127307_t	11	*ETS1*	0.542	1.05(0.77-1.44) p=0.761	1.07(0.70-1.63) p=0.769	1.29(0.95-1.74) p=0.098	1.07(0.77-1.47) p=0.698	0.81(0.53-1.24) p=0.340
rs7146581_c	14	*TRAF3*	0.421	0.86(0.60-1.25) p=0.437	1.08(0.67-1.74) p=0.744	1.40(0.95-2.07) p=0.087	1.14(0.73-1.80) p=0.561	1.18(0.73-1.92) p=0.503
rs7927894_t	11	*C11orf30*	0.608	0.98(0.70-1.38) p=0.913	0.99(0.67-1.48) p=0.977	1.05(0.76-1.44) p=0.767	1.40(0.99-1.99) p=0.061	0.99(0.64-1.53) p=0.971

ALSPAC 4 CLASS

var	Chr	Nearest gene	omni_p	Early-onset persistent	Early-onset resolving	Late-onset resolving
				10.7%	16.5%	10.9%
rs10214237_t	5	*IL7R*	0.983	1.03(0.89-1.19) p=0.695	0.98(0.84-1.13) p=0.749	1.07(0.89-1.27) p=0.489
rs1057258_c	2	*INPP5D*	0.123	1.10(0.92-1.31) p=0.301	0.87(0.74-1.02) p=0.083	**0.81(0.66-0.99) p=0.042**
rs10995251_c	10	*ZNF365*	0.883	1.00(0.88-1.13) p=0.975	1.09(0.96-1.24) p=0.199	0.98(0.83-1.15) p=0.800
rs112111458_a	2	*CD207*	0.246	1.00(0.83-1.20) p=0.991	1.21(0.99-1.47) p=0.057	0.83(0.67-1.03) p=0.087
rs11657987_t	17	*PGS1*	0.382	1.12(0.98-1.27) p=0.099	1.11(0.98-1.26) p=0.100	1.03(0.88-1.21) p=0.693
rs12153855_t	6	*TNXB*	0.056	**1.33(1.06-1.66) p=0.012**	1.16(0.93-1.44) p=0.191	1.33(0.99-1.77) p=0.055
rs12295535_t	11	*PRR5L*	0.129	**1.44(1.04-2.00) p=0.027**	1.24(0.88-1.76) p=0.216	0.64(0.34-1.18) p=0.151
rs16948048_g	17	*ZNF652*	0.682	0.99(0.88-1.13) p=0.925	1.01(0.89-1.15) p=0.890	0.86(0.72-1.02) p=0.080
rs17389644_a	4	*IL21*	0.544	1.11(0.96-1.30) p=0.170	1.13(0.97-1.31) p=0.124	1.01(0.83-1.22) p=0.959
rs17881320_t	17	*STAT3*	0.062	**1.40(1.13-1.74) p=0.002**	1.15(0.91-1.44) p=0.238	1.12(0.84-1.51) p=0.432
rs2041733_t	16	*CLEC16A*	0.912	1.06(0.94-1.21) p=0.333	1.01(0.89-1.15) p=0.904	1.07(0.91-1.26) p=0.394
rs2143950_t	14	*PPP2R3C*	**0.014**	**1.34(1.15-1.56) p=2E-4**	1.01(0.85-1.20) p=0.892	1.13(0.92-1.39) p=0.262
rs2164983_a	19	*ACTL9*	**4E-4**	**1.42(1.21-1.66) p=2E-5**	**1.21(1.01-1.43) p=0.034**	1.19(0.96-1.49) p=0.112
rs2227483_t	12	*IL22*	0.604	1.05(0.92-1.20) p=0.483	0.98(0.87-1.11) p=0.767	1.16(0.99-1.38) p=0.074
rs2228145_c	1	*IL6R*	**5E-4**	**1.32(1.16-1.50) p=2E-5**	0.96(0.84-1.09) p=0.502	0.90(0.77-1.06) p=0.192
rs2897442_c	5	*KIF3A*	0.126	**1.22(1.06-1.40) p=0.005**	0.96(0.83-1.12) p=0.622	0.99(0.82-1.18) p=0.872
rs479844_g	11	*OVOL1*	**0.002**	**1.34(1.17-1.53) p=2E-5**	1.06(0.92-1.21) p=0.434	1.06(0.90-1.24) p=0.510
rs6010620_g	20	*RTEL1*	0.057	**1.24(1.06-1.45) p=0.006**	0.94(0.81-1.08) p=0.381	1.17(0.97-1.42) p=0.094
rs6473227_c	8	*ZBTB10*	0.113	1.09(0.96-1.25) p=0.179	**1.16(1.02-1.31) p=0.026**	1.16(0.98-1.37) p=0.087
rs6602364_g	10	*IL2RA*	0.928	0.98(0.86-1.11) p=0.708	0.97(0.85-1.12) p=0.706	0.91(0.77-1.08) p=0.291
rs7127307_t	11	*ETS1*	0.583	1.08(0.95-1.22) p=0.235	0.97(0.85-1.10) p=0.601	1.13(0.96-1.32) p=0.130
rs7146581_c	14	*TRAF3*	0.364	**1.17(1.00-1.38) p=0.048**	1.10(0.94-1.28) p=0.244	1.07(0.89-1.30) p=0.453
rs7927894_t	11	*C11orf30*	0.132	**1.19(1.05-1.35) p=0.007**	0.98(0.87-1.12) p=0.806	1.11(0.94-1.31) p=0.216

PIAMA 4 CLASS

var	Chr	Nearest gene	omni_p	Early-onset persistent	Early-onset resolving	Late-onset resolving
				6.6%	11.4%	10.2%
rs10214237_t	5	*IL7R*	0.985	1.08(0.79-1.47) p=0.635	0.97(0.71-1.32) p=0.845	1.10(0.82-1.48) p=0.515
rs1057258_c	2	*INPP5D*	0.369	0.79(0.56-1.13) p=0.194	1.12(0.79-1.60) p=0.529	0.72(0.50-1.02) p=0.066
rs10995251_c	10	*ZNF365*	0.876	1.18(0.88-1.59) p=0.278	1.03(0.75-1.42) p=0.853	0.89(0.66-1.20) p=0.443
rs112111458_a	2	*CD207*	0.844	1.18(0.76-1.84) p=0.457	0.83(0.57-1.21) p=0.343	1.19(0.75-1.87) p=0.466
rs11657987_t	17	*PGS1*	0.803	1.03(0.80-1.33) p=0.800	0.83(0.64-1.07) p=0.153	1.06(0.81-1.40) p=0.658
rs12153855_t	6	*TNXB*	0.882	1.32(0.73-2.40) p=0.354	1.23(0.70-2.14) p=0.473	1.19(0.64-2.19) p=0.581
rs12295535_t	11	*PRR5L*	0.557	0.27(0.04-1.69) p=0.163	1.63(0.85-3.14) p=0.144	1.09(0.48-2.51) p=0.836
rs16948048_g	17	*ZNF652*	0.143	1.05(0.79-1.41) p=0.725	**1.40(1.06-1.84) p=0.016**	1.25(0.94-1.66) p=0.124
rs17389644_a	4	*IL21*	**0.031**	**1.72(1.26-2.34) p=6E-4**	1.14(0.83-1.57) p=0.408	1.03(0.69-1.52) p=0.897
rs17881320_t	17	*STAT3*	0.133	**1.62(1.01-2.58) p=0.044**	**1.62(1.06-2.47) p=0.026**	1.10(0.62-1.94) p=0.745
rs2041733_t	16	*CLEC16A*	0.223	**1.37(1.03-1.81) p=0.030**	1.03(0.79-1.36) p=0.818	1.26(0.95-1.67) p=0.114
rs2143950_t	14	*PPP2R3C*	0.517	1.31(0.93-1.86) p=0.124	0.97(0.67-1.41) p=0.878	0.75(0.48-1.15) p=0.189
rs2164983_a	19	*ACTL9*	0.311	1.27(0.86-1.87) p=0.229	0.94(0.63-1.38) p=0.735	**1.49(1.03-2.17) p=0.036**
rs2227483_t	12	*IL22*	0.633	1.19(0.92-1.53) p=0.186	1.01(0.77-1.34) p=0.928	1.25(0.90-1.72) p=0.180
rs2228145_c	1	*IL6R*	0.967	0.96(0.73-1.25) p=0.759	1.12(0.86-1.45) p=0.396	1.04(0.79-1.36) p=0.785
rs2897442_c	5	*KIF3A*	0.798	1.17(0.85-1.61) p=0.332	1.00(0.72-1.39) p=0.995	1.21(0.90-1.64) p=0.215
rs479844_g	11	*OVOL1*	0.457	**1.38(1.01-1.89) p=0.042**	1.05(0.80-1.38) p=0.710	0.92(0.70-1.20) p=0.524
rs6010620_g *	20	*RTEL1*	0.958	1.13(0.81-1.56) p=0.470	1.08(0.80-1.46) p=0.619	1.10(0.78-1.53) p=0.591
rs6602364_g	10	*IL2RA*	0.894	0.86(0.65-1.14) p=0.291	1.11(0.83-1.48) p=0.486	1.00(0.76-1.32) p=1.000
rs7127307_t	11	*ETS1*	0.723	1.10(0.83-1.46) p=0.521	1.09(0.84-1.41) p=0.533	0.80(0.60-1.08) p=0.149
rs7146581_c	14	*TRAF3*	0.627	0.86(0.62-1.18) p=0.344	1.30(0.92-1.84) p=0.131	1.06(0.74-1.51) p=0.764
rs7927894_t	11	*C11orf30*	0.995	1.04(0.77-1.41) p=0.779	1.07(0.81-1.42) p=0.615	1.04(0.77-1.41) p=0.779

Each risk allele is shown after the rs number; p<0.05 shown in bold

* rs6473227 was unavailable in PIAMA